THE UNIVERSITY *of* EDINBURGH

Genetic Evaluation of Traits Recorded in Sport Horses in GB

Isobel Stewart

PhD
The University of Edinburgh
2011

Declaration

I declare that the work within this thesis is entirely my own.

Signed:

Isobel D Stewart

Acknowledgements

I am extremely grateful to my supervisors, Sue Brotherstone and John Woolliams. They have provided expertise, advice, patience and kindness, and without their help this work would not have been possible.

I would also like to thank Ian White for his statistical help and advice. I am also grateful to the WEG group and students in the Genetics & Genomics group at The Roslin Institute for both academic advice and stimulation, and friendship.

Funding was provided by the BBSRC, Biosciences KTN and the British Equestrian Federation and data and information provided by the National Equine Database, British Dressage, British Eventing and British Showjumping. The British Equestrian Federation, also provided collaboration and data for the project, and I am particularly grateful to Jan Rogers and Graham Suggett, who have been involved throughout and who have provided enthusiasm and knowledge from the equine industry.

A special thanks to students and colleagues at Ashworth Laboratories and at The Roslin Institute for their social distraction.

Thanks also to my parents, and siblings for their support in everything.

Publications

The following papers have resulted from this thesis and are included as appendices:

- Stewart ID, Woolliams JA, Brotherstone S (2010). Genetic evaluation of horses for performance in dressage competitions in Great Britain. Livestock Science 128 (1-3), p 36-45

- Submitted to Archives Animal Breeding: Stewart ID, Brotherstone S, White IMS and Woolliams JA. Predicting variance components and breeding values for eventing disciplines and grades in sport horses

- Submitted to Animal: Stewart ID, Woolliams JA and Brotherstone S. Genetic evaluations of traits recorded in British young horse tests

List of Abbreviations

AI	Artificial Insemination
BLUE	Best Linear Unbiased Estimate
BLUP	Best Linear Unbiased Prediction
BS	British Showjumping
BWP	Belgian Warmblood
BE	British Eventing
BD	British Dressage
BEF	British Equestrian Federation
DWB	Danish Warmblood
EBV	Estimated Breeding Values
FEI	Fédération Equestre Internationale
GB	Great Britain
KWPN	Royal Warmblood Studbook of the Netherlands
LRT test	Likelihood Ratio Tests
MCMC	Markov Chain Monte Carlo simulation
MACE	Multi-trait Across Country Evaluation
MAS	Marker Assisted Selection
NED	National Equine Database
PEV	Prediction Error Variance
QTL	Quantitative Trait Loci
RHQT	(Swedish) Riding Horse Quality Tests
REML	Residual Maximum Likelihood
SNP	Single Nucleotide Polymorphism
SWB	Swedish Warmblood
SF	Selle Français
UELN	Universal Equine Life Number
UK	United Kingdom
WBFSH	World Breeding Federation for Sport Horses
YHE	(British) Young Horse Evaluations
YHT	(Swedish) Young Horse Tests

Contents

Abstract

Genetic evaluations for sport horses are performed by many Northern European studbooks, and estimated breeding values - which aid the selection of horses for breeding progeny with good competing ability - are routinely published. Either competition data or young horse evaluation data, or frequently a combination of both is used. Genetic evaluations are not yet performed in GB, and to date the only research conducted has been for the sport of eventing. The aims of this thesis were to: i) perform genetic evaluations for performance in dressage, including considering the effect of breed, ii) perform genetic evaluations for eventing, extending previous work, by using and comparing different methods of producing large (co)variance matrices and using a larger dataset, iii) characterise for the first time data recorded in the recently established young horse evaluation scheme, and if possible perform genetic evaluations and iv) investigate the suitability of international performance data for genetic evaluations for showjumping.

Competition data was obtained from the national competition bodies and young horse evaluation data from the British Equestrian Federation. Pedigree was sourced from both. Appropriate mixed effects models were developed; animal models were used where pedigree data was sufficient, and sire models otherwise. Variance components were estimated using residual maximum likelihood and estimated breeding values were predicted using best linear unbiased prediction. Eventing comprises 3 disciplines, each with 4 grades i.e. 12 traits. For the evaluation, multivariate variance component analysis was performed, using and comparing three different methods of estimating the large 12 by 12 (co)variance matrices required for breeding value prediction. Significant, but low, heritabilities were found for dressage (0.07 – 0.09) and showjumping (0.09 – 0.16), and in the lower grades of cross-country (0.01), indicating that genetic progress for competition performance can be made by selection, particularly in

showjumping. Genetic correlations within dressage and showjumping were high whereas those across disciplines were generally low and only significant for some of the lower grades, particularly between dressage and showjumping. Twelve breeding values were predicted for each horse. The methods based on combining covariance matrices from bivariate analyses were found to be the most reliable. In the dressage evaluations, significant heritabilites of 0.15 without accounting for breed effects and 0.11 when accounting for breed were found. Breed had a significant effect on competition performance, with warmbloods performing better than other breed groups. The model including breed was found to be the most appropriate for genetic evaluations in GB, although not currently implementable in practice due to limited recording of the breed identifier. In the young horse scheme, owners select which discipline (dressage, showjumping or eventing) their horse will enter. Traits are then recorded within the intended discipline. For the analyses, a genetic correlation of 1.0 was assumed between the same trait recorded in different disciplines. Significant, moderate, heritabilites were found for all five traits – veterinary (0.25), type and temperament (0.42), athleticism (0.20), conformation (0.29) and correctness of paces (0.30). The young horse scheme appears to be successful in establishment and uptake. Evaluations were performed for showjumping using international competition data. The small size of the dataset (approximately 6239 records on 479 horses) and large degree of selection limited the study. A larger dataset, including national data, needs to be made available to perform more accurate evaluations. However, an upper limit on the heritability was estimated to be 0.09 (standard error 0.012).

This study explored the use of competition and young horse data for genetic evaluations. The eventing evaluations were the most reliable, and breeding values have been estimated which are ready for publication. However, there are certain limitations to all the competition data sources; there was selection in the data available for dressage and showjumping that biases the analyses.

In addition, there were general limitations in the recording of competing horses, and in pedigree recording that limit the accuracy of the analysis. The young horse data represents a very valuable data source for genetic evaluations, although the scheme is still in its early stages. Ultimately, the most appropriate genetic evaluations will likely use a combination of young horse data and adult competition data. Further work can now perform multi-trait genetic evaluations using multiple data sources, such as combining young horse and competition data, or multiple competition disciplines.

1. Introduction

1.1 The History of the Warmblood

Within the equine sports of dressage, eventing and showjumping, a wide range of horse breeds potentially compete. At the professional and elite levels (and particularly on the continent in Northern Europe), competition is dominated by warmblood horses. Other sport horse breeds such as The Irish Sport Horse, hotbloods (Arabians and Thoroughbreds), and coldbloods (draft breeds), may also compete (the latter in particular at lower and amateur levels of competition). For genetic improvement of the sport horse, therefore, the warmblood is of considerable importance.

Warmbloods originated in Europe, and their history dates back well over 300 years. As the name indicates, they were derived from breeding between the hotblood horses (Thoroughbreds and Arabs) and the coldbloods (draft horses). The many breeds, defined by the studbooks, have unique and varied histories, and most were established from breeding between local warmblood mares and either imported warmblood sires from other breeds or Thoroughbred sires. An illustration of the complexity of the history of the gene pool of warmbloods is given by the Holsteiner breed. This was established in the 1300s, using horses of the Elbe River, Holstein, Germany, which had been bred by monks for war and tournaments. Following the Reformation, the Holsteiner was bred as a coach horse, incorporating genes from the Yorkshire coach horse. After World War II, Thoroughbred, Anglo-Arab and Selle Français (SF) were introduced, to achieve a type suited for elite competitive sport (Celly, 2004). The Holsteiner has contributed to many other warmblood lines, for example in establishing the Hanoverian. Hanoverians have also

been influenced by Thoroughbreds, Trakener and Arabs, which have contributed approximately 35%, 8% and 2.7% of genes respectively (Hamann & Distl, 2008).

In the earlier part of the 20th century, horse populations were largely sustained by their requirements by the military. At the end of WWII this demand was much reduced, and the populations of many horse breeds declined and went through bottlenecks (Fredricson, 2003a). This brought about a loss of genetic variance in the populations. The effective population size of the Hanoverian population born from 1980 – 2000 was estimated at 370 per generation, and the effective number of founder stallions is 364. A few stallions have made particularly significant contributions (Hamann & Distl, 2008).

The warmbloods are now particularly bred for and suited to competitive sport – combining good temperament, medium size and athletic movement. The genetic differences between specific breeds can be slight, given the common founding of one breed from another warmblood breed, and the common introduction of Thoroughbred and mixing of warmblood breeds through history. Some of the modern studbooks – for example those in Germany are distinguished more on geography and market demand than differences in type. And yet, the characteristics of individual breeds, for which they are known and celebrated, highlight their considered distinctions; for example the "intelligence" of the Trakener and the "calm and willing temperament" of the Oldenburg (Celley, 2004).

Contemporary warmblood breeds (assessed using stallions with progeny in young horse tests) are thought to be better connected genetically than between groups of dairy cattle across countries within Aryshire or Holstein breeds (Thorén Hellsten *et al*., 2008; Jorjani, 2000 & Weigel *et al*., 2000 as cited by Thorén Hellsten *et al*., 2008). Ruhlmann *et al*., (2009b) assessed

genetic connectedness between 7 European countries for showjumping stallions; the most inter-connected subset was France, The Netherlands and Germany. The genetic exchange between studbooks has increased over the last 20 years due to greater importation of sires or semen, facilitated by the availability and uptake of artificial insemination (AI) (Thorén Hellsten *et al*., 2008). Most warmblood studbooks are open (in contrast to the closed studbooks of the Thoroughbred and Arabian), i.e. horses of other breeds may be incorporated into the studbook, inline with given criteria, for example those possessing characteristics that will improve performance. One exception is the Trakener, which remains closed, and allows only thoroughbreds to be included (Philipsson, 2007).

The impact of the mixing of breeds was studied in the Swedish Warmblood (Thorén Hellsten *et al*., 2009b). The composition of influence of the foreign breeds has changed with the breeding aims of the horse – i.e. from Hanoverian and Throughbred to the Holsteiner and the Royal Warmblood Studbook of the Netherlands (KWPN). Showjumping performance in the past (before 1979) was shown to have been particularly influenced by The Holsteiner, KWPN and SF, evidenced by higher estimated breeding values for sires of those studbooks. For dressage, the introgression of genes from other studbooks has had a less distinct effect, however the Oldenburg has had an impact recently.

The Thoroughbred, which has made a substantial contribution to many of the warmblood lines, is popularly acclaimed for its intelligence, spirited and bold temperament, speed, agility and stamina. The British Thoroughbred (original) studbook dates back to 1770, and a foundation of 400 horses (Weatherbys, n.d., para.1 & 2). The effective number of founders, thought to have become stable shortly after establishment, is estimated to have been 28. The thoroughbred today is slightly inbred – cumulative inbreeding is 13% (Mahon & Cunningham, 1982), and the

population is essentially closed. The ancestry is particularly well recorded for British horses, which makes it possible to trace back the contributions of specific animals to the population (Cunningham, 2001).

1.2 Warmblood breeding programs

European studbooks breed the great majority of warmbloods, producing 80% of the 130 000 foals registered each year with members of the World Breeding Federation for Sport Horses (WBFSH) (Koenen & Aldridge, 2002). Within Europe, major contributors to foal production are: Germany 30 000 p.a., France 15 000, The Netherlands 12 000, Belgium 4 000, Ireland 4 000, Sweden 3 500 and Denmark 2 500 (Fredricson, 2003b). In Great Britain (GB) the approximate number is 9 000 p.a. (Stewart, 2011), less than The Netherlands.

Breeding programs and selection strategies for enhancing the genetic merit of warmblood breeds are practiced. As well as the traditional improvement of the population through the incorporation of genes of foreign breeds, breeding programs select animals with high phenotypic or genetic merit from within the population for breeding to produce an increase in the mean level of the population. Traditionally breeding programs have been based on selection of phenotypic traits, such as conformation. Subjective assessments and a great deal of faith in ancestry and pedigree are traditional practices and entwined in selection procedures. More objective genetic methods have become assimilated into breeding programs in the latter part of the 20th century, including estimated breeding values (EBVs), which quantify the genetic merit of an animal for breeding for a particular trait (Mrode, 2005). Performance tests, performance rankings supplied by WBFSH and EBVs are now used more or less routinely by individual breeders to select breeding animals (Koenen & Aldridge, 2002).

Breeding programs for European Warmbloods have traditionally been, and remain, more structured than in GB. Stallions and mares with high merit for a particular breeding objective are selected based on one or more of the following selection stages: 1) pedigree, and known requirements of a breeding horse, performed by the breeder, 2) conformation (the correctness of physical body – i.e. body proportions, bone and muscle structure – for the intended purpose, as opposed to features that may predispose to injuries or unfitness for purpose), 3) stallion or mare station or field performance tests to select for breeding by the studbook, 4) separate young horse performance tests, 5) adult or young horse competition results, 6) progeny testing (at approximately 12 years), 7) measures of health or fertility, such as orthopedic status and sperm quality, 8) EBVs, that use either or both of young horse and competition data accumulated from these sources.

The selection intensity for stallions is much greater than for mares. Traditionally, a stallion had the potential to cover a few hundred mares, although typically covered many less. For instance Doruto, who bred from 1965 to 1988 sired 1600 foals (McMahon, 1996, as cited by The Top Dressage Sires of International Horses 1990-1995, 2004). With the availability of AI, this capacity has further increased. Stallion selection strategies are therefore given greater priority compared to mare selection (if the latter is practiced at all). For example, in the SF, for 30 years to 2002, the selection intensities in stallions were nearly four-fold greater than for mares, at 1.95 and 0.48 respectively, whereas the accuracy and generation intervals were similar for both sexes - 0.66 and 12.0 for males and 0.60 and 11.5 for females respectively (Dubois *et al*., 2008).

1.3 Genetic Evaluation Methods

Genetic selection is performed using information on the estimated breeding value of an animal. Breeding values are usually estimated in horse populations using Best Linear Unbiased

Prediction (BLUP). BLUP was introduced into animal genetics in the 1980's, and has since become the primary method of breeding value estimation in livestock. The first application of BLUP in horses was in French riding horses by Langlois (1975) (as cited by Árnason & Van Vleck, 2000); subsequently it has been adopted by many studbooks. The BLUP animal model is the most applied method. To drive the BLUP, variance components are required, which are estimated using methods such as Residual Maximum Likelihood (REML). BLUP is then used to predict EBVs for all horses in the pedigree.

Earlier methods for breeding value estimation included ordinary least squares and generalized least squares. A limitation of least squares methods were that they were particularly suited to balanced data, and that all effects (except for the random error), including animal breeding values, were estimated as fixed effects. The breeding values were then predicted as linear functions of the solutions. Using least squares methods, one of the major problems was that the breeding value estimates with less information were highly influenced by environmental errors, and tended towards the extremes of the distribution, so that the animals with the highest values, most attractive for selection, were often those with little information, such as young sires. As information was added, they would become more accurate but in doing so, those in the upper tails would regress back towards the mean, to be replaced in the tail by others that were less reliable and biased upwards.

BLUP (Henderson, 1973) predicts estimates of random effects, using variance component estimates from REML (Patterson & Thompson, 1971, as cited by Lynch & Walsh, 1998), which are simultaneously and iteratively estimated using random effect estimates from BLUP (Lynch & Walsh, 1998). A mixed model method is used, which simultaneously estimates and accounts for fixed effects (estimating Best Linear Unbiased Estimates; BLUEs) and random effects

(predicting BLUPs). REML enables unbalanced data to be included. The name of BLUP was derived from the fact that it is the best estimate (maximizing the correlation between the true and predicted breeding values), using linear functions of the observations to predict breeding values, and that it is unbiased, in that for breeding values and other random effects, and functions of fixed effects, the expected true estimate is equal to the estimate/ prediction (Mrode, 2005).

BLUP/ REML overcame many of the problems of least squares. It allowed for direct incorporation of different relationships, assortative mating, selection and unequal family sizes, (and unbalanced data of different generations), it was able to estimate non-additive genetic variances (such as dominance, epistatic and maternal, non-additive genetic variance) (a feature of residual maximum likelihood) and used all information on all relatives (a feature of the animal model). One of the main improvements of BLUP, was that breeding value estimates are shrunk back towards the mean, with the degree of shrinkage dependent on the amount of available information, i.e. being greater with less available information. Estimates are therefore less influenced by environmental error, and those with little information are clustered around the mean, rather than being identified as excellent or very bad animals. As more information is added, the estimates go either up or down with equal probability.

For the purpose of presentation to breeders, the EBV horses are usually published scaled to have a mean of 100 and standard deviation of 20. Stallion estimated breeding values traditionally have been produced, although broodmare values have become more frequent (Interstallion, 2005). These are usually published annually in the winter/ autumn.

1.3.1 Application in horse populations

The main breeding objective for most studbooks is performance in the three sport disciplines, conformation and gaits. Koenen *et al*. (2004) surveyed the breeding objectives of 19 European breeding organizations, and their relative importance. Stated objectives were the sport disciplines – showjumping, dressage, eventing and driving - gaits, conformation, health (including soundness, robustness and durability), good behaviour and good fertility. Overall, the greatest importance was placed on showjumping, followed by gaits and then conformation. Many studbooks breed for developing both sports of dressage and showjumping, while a few focus upon one sport discipline, e.g. the Irish Sport Horse and Holstein on showjumping (Janssens, 2008). By comparison, eventing tends to be considered less important, and evaluations are much less common. The market for eventing horses is smaller and they tend to be bred from crosses with Thoroughbreds, rather than using specific breeding strategies (Árnason & Vleck, 2000). Many of the studbooks (e.g. KWPN, Hanoverian, Holsteiner and Swedish Warmblood) produce estimated breeding values for showjumping, dressage and conformation, as separate values.

For some studbooks, despite defined breeding objectives, estimated breeding values are produced for only a selection of these traits. For instance, some studbooks use purely performance data in selection indices, although their breeding objectives include other traits, for instance health. The SF is one example, and produces estimated breeding values for showjumping, dressage and eventing, however not for "conformation, gaits or temper", which are also included in the breeding objective (Janssens, 2008), and include only competition results in the bivariate BLUP model. In part at least, this may be due to a lack of available data, and as this information becomes available, the evaluations may begin to encompass these. In addition they may focus resources on genetic evaluations for traits with the greatest priority or market

demand. Multiple-trait BLUP may use a wider range of traits in selection indices, and therefore may select for other traits indirectly.

Performance EBVs use either young horse performance data, adult competition data or, increasingly, a combination of both. Conformation EBVs use studbook entrance scores (e.g. Holsteiner) or young horse performance data (e.g. Swedish Warmblood) (Janssens, 2008). Multiple data sources, univariate or multivariate models may be used. For instance, for breeding values for each of the sport disciplines, the SF uses a bivariate model using log of annual earnings and ranks of competition performance. The most complex model is that of (some of) the German Warmbloods, e.g. Hanoverian, which produces a combined estimated breeding value of the performance traits using a multiple trait, repeatability animal model with 15 traits including both adult and young horse competition performance in showjumping and dressage, and results from stallions' and mares' tests (Janssens, 2008). If multiple estimated breeding values, for different traits are predicted for each animal, these may then be presented individually, or as an index (Koenen & Aldridge, 2002) with values weighted to achieve a combination of breeding objectives.

Other factors taken into account in the evaluation include sex, age, place and time, permanent environment and some grading of rider (Janssens, 2008). Stallions and geldings tend to perform better than mares (Koenen & Aldridge, 2002). The rider potentially accounts for a relatively large amount of variance (18%, 10% and 15% in the advanced grade of eventing in GB, for the dressage, showjumping and cross-country phases respectively) (Kearsley *et al.*, 2008), and should be accounted for if possible; however is frequently not included (Janssens, 2008) due to lack of recording. The environmental effect of stud farm could potentially be included in the model in countries where large stud farms are common.

Few models account for differences between breeds (Thorén Hellsten *et al*., 2009b; Janssens, 2008). However, the impact of Thoroughbreds is claimed to be underestimated in BLUP evaluations within the industry (Thorén Hellsten *et al*., 2009b) and some models currently take percentage Thoroughbred into account in the model (Janssens, 2008). Most evaluations are within studbooks, which place restrictions on entry criteria so that the population is relatively genetically homogeneous. When including multiple breeds in a national evaluation, one of the consequences of not including genetic groups is that EBVs may be regressed toward the national reference population rather than their own, foreign, population, and so may result in an over or under estimation of the EBV (Phocas & Laloë, 2004). The severity may be dependent on the relative numbers of each breed, defining the proportion of foreign genes that make up the gene pool, and can be overcome by fitting genetic group in the model. The effect of including genetic groups was examined for the Swedish Warmblood. The authors found that the advantages were not sufficiently great to warrant it – as pedigrees with depth were available, and the breed is a mix of other warmblood breeds. Correlations between EBVs estimated with and without genetic groups were 0.99, and EBVs were less accurate when estimated with genetic groups. Assigning foreign horses to genetic groups was also problematic given the high degree of mixing between populations (Thorén Hellsten *et al*., 2009b). However, in a population including a wider range of different breeds, or with less extensive pedigree, the inclusion of genetic groups in the model or an across-breed evaluation is likely to be advantageous. Across breed evaluations are becoming more common for dairy cattle.

1.4 Genetic Progress

The genetic gain (R) achieved is dependent on the selection intensity (i), heritability (h^2), additive genetic variance (σ_a^2) and generation interval (L) in the specific population as $R = i\ h\ \sigma_a / L$. The selection intensity, i, is dependent on the proportion of the population selected, and

indicates the difference in the mean of the selected individuals from the population mean (i.e. the superiority), measured in standard deviations.

There is evidence for genetic progress in warmblood populations, particularly over the past 30 years. Evidence for progress prior to this suggests it was limited, possibly due to a lack of active selection of competition horses (although in some populations this may be difficult to assess due to lack of available data). More recently, the use of BLUP, AI, and increased international exchange of genetic material are likely to have enhanced the rate of genetic progress. For instance, in the Swedish Warmblood (SWB) from 1980 to early 2000's the genetic progress for showjumping was estimated as 0.93 genetic standard deviations, and for dressage 0.50 genetic standard deviations (Thoren Hellsten *et al.*, 2009b). Prior to this (1973 – 1979) there was no detected genetic progress for conformation, and moderate progress for gaits and jumping (Árnason, 1987). In the SF genetic progress was 0.056 and 0.096 genetic standard deviations per year for showjumping from 1985 – 1995 and 1995 – 2002 respectively (Dubois & Ricard, 2007).

The most rapid genetic gain would be achieved through (multivariate) selection for one breeding objective, rather than for multiple traits. In multi-trait selection the effectiveness relies on the magnitude and direction of genetic correlations between the traits. If a negative genetic correlation exists between two traits, then it can be difficult to achieve genetic gain in both traits, whereas if a positive correlation exists, then it is much simpler to make positive gain in both traits. Most studbooks select for dressage and showjumping. Current evidence as to the genetic correlations between these is unclear – some studies suggest weak positive correlations, others suggest negative correlations (Árnason & Vleck, 2000; Huizinga & Vandermeij, 1989). The majority of evidence is from correlations between traits recorded at young horse tests, where dressage and showjumping related traits are weakly and sometimes negatively correlated.

Slightly higher correlations are observed between canter and the jumping traits as opposed to walk or trot (Thorén Hellsten *et al*., 2006). Eventing evaluations have provided some evidence for genetic correlations for adult competition performance. Low/ moderate correlations were observed between all disciplines (showjumping, dressage and cross-country) at the lowest level of competition in GB, although these estimates were lower and had a greater degree of uncertainty at higher levels of competition (Kearsley *et al*., 2008). If this reduction in correlation were confirmed, the correlations at lower levels may reflect a general athletic ability and good temperament as opposed to correlations between specific jumping or dressage related traits.

An important question is whether specialist showjumping and dressage horses, suited to competition at Olympic standards is the key objective, or more all-round horses. The level of competition for which horses are intended is important, as some traits may be of differing importance at the different levels – for example for horses intended for leisure use, the primary importance may be on good temperament as opposed to excellence in jumping.

In multi-trait BLUP selection for a single trait, the response will always be greater compared to univariate BLUP. A simple way of assessing the potential value of including an indirect trait is that the value will be greater if $h_2 r_A / L_2$ is large relative to h_1 / L_1, where h_1^2 is the heritability of the selected trait, h_2^2 is the heritability of the correlated trait, r_A is the genetic correlation, and L_1 and L_2 are the generation intervals of the selected and correlated traits respectively (extending the analysis of Falconer and Mackay, 1996). This may be of particular interest in considering the advantages of young horse traits versus adult competition data. Often the breeding objective is for adult competition performance, however, the high generation interval makes the use of indirect selection, on traits other than competition which are available at a younger age, attractive.

1.5 Young Horse Data

Young horse testing is primarily conducted for the purpose of selecting stallions and mares for breeding and registration with the studbook, for collecting data for genetic evaluations, and for selecting animals for competing later in life. Specifically designed stallion or mare selection tests are used to select animals with high breeding merit as judged by the breeding organisation (in addition the data may also be used for genetic evaluations and to identify horses with high competition potential). From these tests, most horses go on directly to breed, while some compete. These tests are conducted either at field or at station at 3 to 4 years of age (illustrating the shorter generation interval compared to adult competition data). Station tests, primarily conducted for stallions, are of long duration (commonly stallion tests are 70 days). Station tests for mares are also performed in some countries. Field tests are aimed at mares and geldings, although a few studbooks test stallions in this ways, and are of a much shorter duration (usually 1 day) (Thorén Hellsten *et al*., 2006). Other tests include ordinary young horse performance testing or competitions, which aim to identify horses with high merit for competing later in life, and to enable genetic evaluations of these horses and their parents. Tests are conducted between 3 to 6 years of age. Traits such as the gaits (walk, trot & canter), rideability and jumping ability are examined in young horse tests (Thorén Hellsten *et al*., 2006). Competitions for young horses are common in many countries (Thorén Hellsten *et al*., 2006), for example, in France the "Cycle Classique" is an age-grouped competition in all disciplines (primarily showjumping, but also dressage and eventing). These are used for stallion and mare selection e.g. in France and Belgium (Ricard *et al*., 2000), as well as to identify horses for competition.

There are various advantages and disadvantages to the use of data from station or field tests in genetic evaluations. The station test is performed in a controlled environment, and therefore

subject to less environmental variance. The repeatability is high, as are heritabiltities and correlations with later competition performance (Lüehrs-Behnke *et al*., 2006a). However, the cost and resources impose major limits to the number of horses that can be tested. Station test data is used for evaluations by, for example, the studbooks of Denmark, Germany and The Netherlands (Thorén Hellsten *et al* 2006; Koenen & Aldridge, 2002). Field tests are conducted in a much larger environment, with greater environmental variance. A large number of horses can be examined (for example 60% of mares of the Hanoverian studbook are tested in this way), and due to this larger testing capacity field tests are subject to less pre-selection. However, they are less accurate, subject to greater biases, and have lower heritabilities and correlations with later competition performance. The Swedish use data from a 1 day field test for genetic evaluations (Olsson, 2000). Other countries that use field test data include Norway, Finland, Denmark, Germany and The Netherlands (Koenen, 2002; Thorén Hellsten *et al*., 2006) and the in the United Kingdom (UK) (Stewart, 2011). Denmark, Germany and The Netherlands use data from both field and station tests (Thorén Hellsten *et al*., 2006). The relative merits of the use of station or field test data have been examined in some populations. In both Swedish and French populations, the importance of large-scale testing, available in field tests has been noted (Philipsson *et al*., 1990; Dubois *et al*., 2008).

Young horse data may only cover a small proportion, and likely biased, selection of animals, as horses may never participate in young horse tests, particularly for stallion and mare tests. These will reduce estimates of genetic variance derived from evaluations, and so heritability estimates. Young horse data may be subject to biases with judging and judging procedures. Changes over time in the data collection potentially also introduces biases, for instance due to differences in traits or recording patterns. This was observed within the Swedish data so that genetic

correlations between traits recorded in the early and later years of a thirty year period ranged from 0.48 to 0.97 (Viklund *et al.*, 2008).

A major advantage of young horse test systems is that they are quick to establish, so now some countries have datasets covering the last 10-20 years. Examples include The Netherlands, Sweden, Switzerland and Germany (Ricard *et al.*, 2000; Thorén Hellsten *et al.*, 2006). A young horse testing scheme was established in the UK in 2005 (in its current form). Also there is a short generation interval from birth to obtaining the data (3-6 years), which allows selection to be practiced with short generation intervals, so achieving a greater response in a given time. For young horse traits to be of value in selection for later competition success, a prerequisite is moderate to high heritabilities and correlations with adult competition performance. Depending on the magnitude of the heritabilities and genetic correlations between young horse and adult data, as well as the respective generation intervals until young horse and adult competition results become available, selection using young horse data can make a major contribution toward achieving a genetic response (whether alone, or included in multi-trait BLUPs).

1.6 Competition data

Adult competition results are recorded by the national regulatory discipline bodies or by the breeding bodies. The nature of the competition data recorded will vary with national practice. For instance, the (ability) levels that are recorded, the proportion of competitors from a competition (i.e. some countries record all competitors and their placings, whereas others will record just a selection of horses finishing in the highest places) and the performance measures (e.g. placing or earnings) may all vary. Horses performing in “adult” competitions may begin their competitive life at four years (so overlapping in age with young horse tests).

The measure of performance varies, and includes the points awarded, penalty points awarded, ranks, earnings, or highest level achieved which may be used as an individual competition measure or as an aggregate measure e.g. annual or lifetime. These often require transformation before analysis with linear models. In the case of earnings, this is particularly apparent, as the nature and incentive of monetary awards is that the increments between sums increases with the place.

Many countries have recorded competition data in a relatively consistent format for a long period of time. Therefore, there exist large datasets containing data related to a high number of horses. Competition data is likely to cover a large proportion of the horse population, as many horses will compete at some point in their lifetime. Selection will increase at higher ages and levels of competition, as only those capable of doing well will continue to compete. A small proportion of horses may never compete, due to a lack of ability. A further proportion will never compete due to exclusive use for breeding; whilst some may be removed from competitive careers for periods to use for breeding. The system used to record results may introduce further selection, as only a selection of results - e.g. those scoring points - may be recorded depending on the system rules. Some national bodies record only the highest placed or earning horses. Also, only competitions at a particular minimum standard may be recorded. The extent to which this selection occurs differs between countries/ recording systems.

Competition data may be subject to biases with judging and judging procedures, for instance differences in the judging procedures between venues of competition. Some evidence for this may be seen in Dressage competitions in GB, where the environmental effects of class-event interactions, potentially due to judging effects, accounted for 0.100 (standard error 0.0029) of the phenotypic variance. Changes over time in the sport may have occurred (for example due to

differences in recording patterns/ competitions over time and trait definitions). Some evidence for possible changes were found in Swedish data, where genetic correlations between competition traits in the 1950's to 80s with traits at the turn of this century was 0.40 – 0.85 for dressage (the range depending on the performance measure – points per placing and points) and 0.54 – 0.71 for showjumping respectively (estimated using multi-trait animal models) (Viklund *et al*., 2010).

1.7 Trait Definition

The breeding goals of the studbooks testify to the fact that selection traits are often difficult to define, let alone quantify. For example the Oldenburg breeding objective is "A noble, generously lined, high performing sport horse with active impulsion and space gaining, elastic movements which, because of its predisposition, is permanently suitable for any type of sport (dressage, showjumping, eventing, driving)" and the Trakener's is "Sound horse with Trakehner type, a big frame and correct and harmonious body proportions. Versatile riding and performance horse, easy to ride, with an energetic, elastic and ground-covering way of moving. Good and stable temperament. Spirited but kind, intelligent, very willing to perform and with a tremendous stamina" (Janssens, 2008). The problem with these objectives is that they lack clear definition of traits. In general, many objectives are subjective (e.g. noble, kind), and precise definitions are often lacking both within studbooks, and between studbooks (e.g. correct body proportions), leading to a lack of consistency and transparency. The conformation definitions vary between studbooks. Subjective and aesthetic terms are common. On any one trait, the selection intensity is reduced therefore some objective measure of the importance of objectives is required, for instance, using an economic or desired gains approach. At present, the optimum number of objectives, their relative importance and genetic correlations are not known/ clear, which leads to less favourable genetic gain (Koenen *et al*., 2004).

While some performance, conformation and health measures may be relatively easy to obtain measures of, for others such as temperament, soundness, longevity, health and behaviour there are greater problems. For example, soundness is a major breeding objective for the British sport horse, however difficult to quantify. Similarly, temperament is of high importance in selection for performance (Olsson, 2006) however objective testing is difficult. Some traits are used as, or assessed by indirect measures, for instance aspects of conformation are used as indicators for health, longevity and adult performance (Koenen *et al*., 2004). In comparison to performance, the data on health is very sparse and very limited. Restricted to a very small group of horses, it is however conducted for some stallion station tests prior to entry into the studbook, which might examine for example cryptorchidism, bone disease, genetic defects and roaring (interference in air flow in respiration). Until recently, the only studbook that performed more extensive health tests was the SWB, who examine orthopedic and medical status at young horse performance testing of 4 year olds (Wallin *et al*., 2001; Koenen & Aldridge, 2002; Thorén Hellsten *et al*., 2006).

Alternative sources of information could provide valuable traits for evaluations. For instance, fertility recording is performed by semen quality tests of young stallions, and some studbooks publish foaling percentages (Koenen & Aldridge, 2002). Longevity data is not recorded as standard, although conformation and length of competitive life could be used as indirect measures from available data (Wallin *et al*., 2000; Ricard & Fournet-Hanocq, 1997). Before alternative data sources are used, however, the genetic correlations between the selection and indirect traits would preferably be characterized, for instance for aspects of conformation as a measure of longevity.

At present, these problems in definition and measurement and lack of genetic evaluation inhibit the direct inclusion of many of traits in routine breeding systems. Better measures are required, as well as genetic parameters and correlations with performance. A positive genetic correlation between these desirable traits and competition performance is preferable; if a negative correlation exists, selection will be more difficult.

1.8 International comparisons

For dairy cattle, international estimated breeding values are currently routinely predicted and used in participating countries as breeding tools. Between 20 and 30 countries participate, depending on the breed and trait and estimated breeding values are produced 3 times each year. In horse populations, international comparisons of EBVs, and breeding schemes have been made particularly relevant over the past 20 years, due to increased trade of stallions across countries, and the availability and uptake of AI (Thorén Hellsten *et al*., 2008). This has resulted in increased genetic exchange of material, and mixing and homogenization world wide. Countries with a large population for breeding - for instance France and Germany - have become big exporters, and the genes of their pedigree is now highly represented and introgressed into populations world-wide (Koenen & Aldridge, 2002). Genetic connectedness between a number of European populations (SWB, Danish Warmblood (DWB), KWPN, Hanoverian and Holsteiner), as assessed using young horse performance data has increased over time, with the German studbooks shown to have been the major exporters (Thorén Hellsten *et al*., 2008).

At present, national genetic evaluations are not directly comparable. A stallion may have EBVs published in two or more different countries, where the EBVs differ due to the fact that the data, data recording systems, traits and trait definitions used for evaluations may differ (Koenen & Aldridge, 2002), the definition of the genetic base may differ and different transformation

procedures may be used (Interstallion, 2005). To make evaluations more comparable, harmonisation of the reference population and publication scales have been suggested. The recommendations are to standardize the publication scale to a mean of 100 and standard deviation of 20, using the genetic standard deviation. The reference population (used to define the mean) is recommended to be all stallions with 15 tested progeny, born within a given time period prior to the time of evaluation (i.e. a moving reference population), where the time period is chosen to give a mean of the active breeding population of 100, i.e. 4 – 18 years (Interstallion, 2005). At present, most organizations use a mean of 100 and standard deviation of 20 for the publication scale, with exceptions being the French and Dutch. However, the reference populations differ, either being all stallions or all tested horses, potentially with age restrictions, which results in further differences between countries in the genetic level of the mean. National EBV estimates may include foreign sires, but with EBVs estimated with less accuracy, without fully accounting for the genetic level of the foreign population, particularly for those with weak genetic ties (Phocas & Laloë, 2004; Thorén Hellsten *et al.,* 2009b).

A working body, "Interstallion" was established to explore some of these issues, including "exploring ways of harmonising and comparing EBVs across countries" (Ruhlmann *et al*., 2009a). Joint evaluations rely on sufficient genetic connectedness between populations in order to estimate differences in the mean genetic level of the population and genetic correlations between traits. Genetic connectedness between groups of European countries using young horse and showjumping data were estimated, and found to be sufficient for the estimation of genetic correlations (Thorén Hellsten *et al.,* 2008; Ruhlmann *et al.,* 2009b). Genetic correlations for young horse traits were estimated between the Swedish and Dutch Warmblood using Multi-trait Across Country Evaluation (MACE) (Thorén Hellsten *et al*., 2009a). Correlations were very high for showjumping related traits (0.99), and for dressage traits (0.89–0.97), indicating that

(largely) traits by the same definitions are being selected for. Ruhlmann *et al*., (2009b) estimated genetic correlations for showjumping competition data between Sweden, Denmark, France, Belgium and Ireland. Genetic correlations were high (0.86+) or relatively high (0.70+), depending on the countries. Overall, this evidence suggests that joint genetic evaluations including a number of countries (using MACE) may become feasible for performance traits.

In addition to these scientific issues, there are more practical issues that make international comparisons difficult. A major problem is a lack of a unique and consistent identifier for horses, so that a horse may at present have multiple different identifiers in different countries (as recorded by studbooks or competition bodies). The Universal Equine Life Number (UELN) was introduced in 2004, to be phased in gradually internationally (http://www.ueln.net/welcome/). It is now mandatory in GB, and recorded in horse passports. This will act as an international, unique identification which will be a great benefit to future identification of horses. Another, and associated, issue is the availability of pedigree data, where again, there may be problems in identification of horses. In addition, this may be recorded to different extents. The situation at present is that pedigree data is usually recorded by the studbook and is frequently required for registration. Pedigree records may be very complete and go back many generations, particularly in the case of closed studbooks where entry to the studbook is strictly restricted based on pedigree. In other studbooks, the recording and legislation relating to pedigree may be less complete. This may result in differences in the extent and quality of pedigree data available. The accuracy of EBV estimates will be reduced with incomplete pedigree. This is an issue that affects both national and in some evaluations particularly foreign stallions. The latter was illustrated in a Swedish study, when foreign stallions were simulated with no pedigree available, the average EBV changed by approximately one genetic standard deviation (Thorén Hellsten *et al*., 2009b).

One of the necessary steps for utilizing international evaluations is the free but managed exchange of national data. In horses, progress towards this appears to have been difficult, which has restricted achieving genetic progress in these populations.

1.9 Genetic evaluations in the UK

A main aim of the current breeding initiatives in GB is to perform genetic evaluations and to produce and publish estimated breeding values. In GB, genetic evaluations have been performed for research purposes for eventing (Kearsley *et al.*, 2008) and dressage (Stewart *et al.*, 2010). Estimated breeding values are not published. With these aims, the objective of my PhD thesis is to lay the foundations for genetic evaluations in GB. This thesis has the objective of going beyond the work of Kearsley (2007), which was restricted to performing work on eventing, towards giving a comprehensive suite of models and analyses toward genetic evaluation in GB.

In chapter 2 I will perform a genetic analysis of dressage competition data, for the first time in GB data, which includes a competing population of horses of different breeds. I will analyse the data twice – including and excluding breed from the model - and discuss the advantages of each method. In Chapter 3, I will advance the analysis of eventing data beyond the work of Kearsley (2007). In GB, eventing has four grades of competition for each of the 3 disciplines, thus giving 12 traits for the combinations of each discipline within each grade. Estimating variance components and predicting BLUPs for such a large number of traits presents challenges for computing resources. The aims of the work will be to investigate new methods of producing large (co)variance matrices and then to predict 12 breeding values for each animal, producing the material for an adaptable index depending on the breeding goal. In chapter 4, I will analyse the Futurity data – young horse evaluations introduced by British Breeding in 2005, and perform a genetic analysis for the first time. In chapter 5, I will analyse showjumping data in GB for the

first time, using data collected by the British Showjumping Association. In chapter 6 I will discuss the outcomes of the work in a wider context.

2. Genetic Evaluation of Horses for Performance in Dressage Competitions in the GB

2.1 Abstract

Genetic evaluation of sport horses for dressage competition and estimation of breeding values are performed in many Northern European countries. To date, no such system has been used in GB. The aim of this study was to estimate genetic parameters for performance in dressage competition and then to predict breeding values for horses competing in dressage in GB using an animal model. The percentage of marks awarded was used as the performance measure. Random effects comprised additive genetic, horse permanent environment, class, event and class–event interaction. Horse gender, age, height and competition standard were included as fixed effects. In a second model, breed class was added as a fixed effect. Heritability (standard error) of performance in the two models was 0.15 (0.018) (no breed) and 0.11 (0.036) (including breed). In both models age was significantly associated with performance (P<0.001), with performance peaking at age 10. Stallions performed significantly better than geldings. A significant association between performance and height was detected only in the model which did not include breed effects. The model including breed was considered to be the most appropriate, given the nature of the sport horse population in GB. It should be possible to predict breeding values of sufficient accuracy.

2.2 Introduction

Genetic evaluation of sport horses for adult dressage competition is performed in many countries. The approach can take different forms — either direct evaluation based on adult

competition results, or using an indirect measure of performance from traits examined at young horse tests, or a combination of both. Most evaluations now use a combination of young horse and adult data, including the SWB (Thorén Hellsten *et al.*, 2009a), KWPN, Hanoverian, Oldenburg, and Trakenner (Janssens, 2008; Lüehrs-Behnke *et al.*, 2002). Estimated breeding values from many of these evaluations are routinely published, to inform and aid selection of sport horses for breeding to produce progeny with high performance ability. Success in breeding elite internationally competitive horses is evident among these studbooks, as demonstrated by the high representation of horses from these studbooks competing in the dressage event at the 2008 Olympic Games, with Hanoverian, KWPN, Oldenburg, and DWB being the most represented (FEI, 2008).

Dressage tests the gaits, ridability, athleticism and transition between movements of the horse under the rider. Horse and rider complete a series of predefined movements (a test) in an arena of a specific size which is scored by expert judges. Genetic evaluations based on performance in dressage competition use a variety of measures such as the rank (DWB, SF), highest level (KWPN) or earnings (SF), derived from an individual competition, or from annual or lifetime cumulative performances. The distribution of these measures is often inappropriate for analysis with linear models without transformations to normalize the distribution and make variances homogeneous between competitions/groups. Most protocols for routine competition evaluations exclude the highest levels of competition, such as The Olympic Games and World Championships (Koenen, 2002).

There are advantages and disadvantages in using either adult competition results or young horse tests in evaluations. Competition data typically covers a large proportion of the population, even though only a selection of competition results may be recorded, whereas young horse data may

be highly selected. There is a relatively short time period from birth to obtaining the data (3–6 years) in young horse tests compared to competition results, promoting selection with shorter generation intervals, which in turn gives more opportunity for generating genetic gain. However young horse traits will only be effective in selecting for later competition success if they are heritable and genetically correlated with later adult competition performance. An important practical consideration is the number of records available since this, along with the heritability, will determine the accuracy of prediction. British Dressage (BD), the governing body for the sport in Great Britain (GB), has collected data on competition results since 1978, while the young horse performance tests were introduced by the British Equestrian Federation (BEF) as recently as 2002, and are still developing. Therefore while the optimum system of genetic evaluations will make use of both sources, the priority in GB for introducing evaluations is to use competition data.

Most evaluations are conducted by studbooks and therefore refer to a relatively homogeneous gene pool, particularly for those that have remained closed for a long time period, and few evaluations include a number of breeds. An exception is the German national evaluation, which includes all Warmblood breeds. However breed is not explicitly included as an effect in the model and all the breeds are regarded as being sufficiently related to treat as a homogeneous group. In contrast, the GB dressage data relates to a large number of different breeds, including Warmbloods and native British ponies. Therefore evaluations based on GB competition data will need to encompass many breeds, and hence will differ from the single breed evaluations produced in other countries.

In general, the same methodology is used to perform routine evaluations across all countries (Koenen, 2002). Typically, mixed effects models, using REML are used to estimate variance

components for random effects (e.g. additive genetic and horse permanent environment), while simultaneously assessing the effect of fixed variables on the horse's performance. After estimating these parameters BLUP animal models are used to compute estimated breeding values for all animals in the pedigree. A pre-requisite for this analysis is sufficient genetic connectedness within the population, which is a function of the relationships between the horses (e.g. half-sibs, cousins) with records, and therefore an important aspect of developing genetic evaluations is accumulating the pedigree within the databases and ensuring its integrity.

No genetic evaluations for dressage performance have yet been developed in GB, excluding the dressage phase of eventing (Kearsley *et al*., 2008), and this study has the objective of developing appropriate models that can be used for routine calculation. It is also intended that the availability of predicted breeding values to horse breeders and owners should be as wide as possible. For these reasons, as argued above, development of the evaluation will need to be based initially upon competition data collected by BD. This is the first use of this data for genetic evaluations and the study needs to address the integrity of pedigree recorded, the options for traits upon which to base predicted breeding values, and the development of a parameterized model across multiple breeds.

2.3 Materials and Methods

2.3.1 Structure of Dressage Competitions in GB

BD is responsible for regulating and recording results of dressage competitions in GB, and for recording results of British riders/horses competing internationally. In a competition, the horse and rider receive marks for completing a specific predefined series of movements, known as a test, with higher marks indicating a better quality of performance. A class is the term given to a particular type of test with further definitions, for example class E48SRQ refers to test 48 (which

is elementary standard) performed as a summer restricted- qualifier. A class is conducted over multiple locations and dates, and a specific location and date will be termed an event. At each event, a number of different classes will be run. Classes are grouped into standards, so that standards are an attribute of the test rather than the horse. There are 12 competition standards in the BD data, among which 5 display explicitly a grading in ability (in decreasing order: Advanced, Advanced Medium, Medium, Elementary and Novice), 3 are concerned with young riders (Under-21 Young Riders, Under-21 Juniors and Under-21 Ponies), with 4 others (Freestyle, Area Festival, Medium Elite and Horse Trials).

2.3.2 Databases and data structure

National and international competition results from 1978 onwards were obtained from BD including all BD affiliated competitions. This was combined with pedigree data obtained from the National Equine Database (NED) which contains records for all horses with a passport and pedigree data for some studbooks. All horses born or resident in the UK on or after 2004 are required to have a passport. Pedigree data for some horses preceding the introduction of passports were available. The initial dataset was formed including all competition records since 1978 that satisfied the following criteria: i) the horse was registered to a studbook with pedigree data available in NED, and ii) as a general rule (although not completely adhered to) the horse had scored points for performance, which corresponds to a percentage mark of 60% or greater. BD routinely only record competition results in the database where the horse scores points i.e. gains 60% or more. Note that although BD only collects data on GB horses, i.e. excluding Northern Ireland, NED covers all of the UK, and so EBVs from Northern Irish horses may be predicted from the analysis.

2.3.3 Data handling

Initially, competition records without details of the event, age or gender of the horse, or where both parents were unknown were discarded (leaving 183,551 from an original 222,369 records). Data was then cleaned to address a number of quality issues. Firstly a simplified model for the classes and standards was addressed by reducing standards to the 5 that were indicative of progression: Advanced, Advanced Medium, Medium, Elementary and Novice. These had contributed the vast majority (98.8%) of records. Secondly, the culled dataset had a relatively low number of mares (contributing 1.4% of the data), and further investigation showed that the ratio of mares to the other genders was lower than the corresponding ratios for horses registered with BD. The cause of this was unknown, but may have been due to a lack of pedigree recording for mares. As a precaution to ensure integrity of the dataset, all mares were excluded.

Table 2.1. Number / percentage of records by standard, for each model and within Model 2 breed groups.

		Number/ *percentage* of records by standard						
		A	**AM**	**M**	**E**	**N**	**Total records**	**Number horses**
Model 1		8478	8473	24365	40484	62709	**144509**	**6622**
Model 2		2837	2658	8247	14515	23134	**51391**	**2183**
Model 2 breed groups	**BN**	*0.5*	*1.9*	*9.0*	*27.5*	*61.0*	**4114**	**226**
	Irish	*0.8*	*1.1*	*9.1*	*31.7*	*57.3*	**4983**	**291**
	NBN	*3.7*	*4.8*	*12.5*	*24.7*	*54.4*	**546**	**39**
	TB	*4.9*	*4.7*	*14.3*	*27.5*	*48.6*	**5339**	**230**
	WB	*6.9*	*6.2*	*18.2*	*28.0*	*40.6*	**36127**	**1372**
	Arab	*0.0*	*1.1*	*2.8*	*26.2*	*69.9*	**282**	**25**

Standards are Advanced (A), Advanced Medium (AM), Medium (M), Elementary (E) and Novice (N).
BN = British native/ indigenous, Irish= Irish (excluding thoroughbreds), NBN = Non-British native indigenous, TB = Thoroughbred (including Irish), WB = Warmblood/sport horse, Arab = Arab

The pedigree data were a further issue for cleaning. Names, rather than unique horse identifiers, were recorded in the BD database, with errors in data entry. Therefore it was often difficult to distinguish whether animals were distinct or the same when they had similar or common names. There was no clear, objective solution and so a process of manual cleaning was carried out by a series of steps. This included internet searches on ambiguous entries and consideration that incorrect information has two times the detrimental effect on reliability of genetic evaluation as missing information (Sanders *et al*., 2006; Woolliams, 2006), so that changes erred on the side of caution. In the data analysed, first generation pedigree data were available and largely complete; 99.6% of competing horses (n=6622) had sires and 86.5 % had dams, where the total number of sires and dams were 3802 and 5303 respectively.

Various performance measures were recorded on subsets of the data, including the marks awarded, the percentage of marks of the total available for the test, the points awarded based on the percentage, the placing and the lifetime best. Place could not be used for analysis, as the total number of competitors was unknown. Points were a crude measure based on the percentage awarded, and marks were an unstandardised version of percentage. Lifetime best was a very crude measure, and biased, in that horses were at different stages of their competitive careers. The percentage was available for all competitions scored using the percentage mark method (since 1994), and as it constituted 81% of the cleaned data, including most recent data, percentage was chosen as the performance measure. After all cleaning, 144,509 records remained for analysis.

Breed data was available for only 36% (51,391) of records. Although more information results in more accurate EBVs, breed may have a significant influence on performance, and therefore it was of interest to examine the effect of breed based on a subset of data. However this model

cannot be implemented in practice at present due to the large number of horses with no breed recorded.

2.3.4 Analysis

Mixed effects models, using REML and the program ASReml (Gilmour *et al.*, 2006) were used to estimate variance components for random effects (additive genetic, permanent environment of the horse, class, event and class by event interaction), while simultaneously fitting the effect of fixed variables on the horse's performance. Note that 'event' is defined by both date and place. The fixed effects were gender (1 d.f.; stallion or gelding), and competition standard (4 d.f.), with covariates for age on the day of the competition and height fitted as cubic and quadratic polynomials respectively (the order of polynomials was based on a preliminary analyses). An animal model was implemented, which includes the relationships between the animals. However in the data, relationships involving common grandparents or more distant ancestors were only known if a sire or dam also competed and its parents were known.

The first animal model (Model 1) was fitted to 144,509 competition records:

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \Sigma_{r=1}^{3}\gamma_r x_{ijklm}^r + \Sigma_{r=1}^{2}\delta_r z_m^r + u_k + v_l + w_{kl} + a_m + c_m + e_{ijklm}$$

where y_{ijklm} is the percentage score for the horse m of gender i competing at standard j, in class k of event l. The fixed terms are: μ the mean score, α_i the effect of gender i, β_j the effect of standard j, γ_r the regression coefficient of degree r for the polynomial describing the effect of age x_{ijklm} at the time of the competition, and δ_r the regression coefficient of degree r for the polynomial describing the effect of height z_m. The random terms are: u_k the effect of class k, v_l the effect of event l, w_{kl} their interaction, a_m and c_m representing the estimated breeding value

and the effect of permanent environment for horse *m*, and e_{ijklm} the residual error. The significance of random effects was examined using an approximate t-test, and where the variance component was obviously significantly greater than zero, no further test was done. A likelihood ratio test was performed for those variance components where the significance was less obvious. The significance of fixed effects was examined using an approximate F-test. Differences between levels of fixed effects were assessed using the t-test.

The random effects are assumed to have normal distributions with mean 0. In particular V_p, the total phenotypic variance, was taken as the sum of variance components for all random effects in the model, i.e. including components for class, event and class–event interaction. Heritability (h^2) and repeatability (r^2) estimates were calculated as $h^2=V_a/V_p$ and $r^2=(V_a+V_c)/V_p$ where V_a and V_c are the additive genetic and permanent environmental variance respectively.

In a second model (Model 2), breed classes were added as a fixed effect. Six classes were included: i) British native/ indigenous; ii) Irish, excluding those known to be thoroughbred; iii) Non-British native/ indigenous; iv) Thoroughbred, from anywhere in the world; v) Warmblood/ sport horse; and vi) Arab. This classification followed the advice of British Breeding (Table 2.1). Note that these classifications are likely to contain cross-bred animals; however the extent of this cannot be determined. This model was fitted to 51,391 records after excluding horses without breed. In addition, to enable a comparison of the breed model with a non-breed model, the analysis was re-run on the dataset that included breed, but without including breed as a term in the model. EBVs from both models were obtained (the breed effect was included for Model 2), to examine their distribution, the genetic trend over time based on year of birth for all competing horses, and the correlation between the two models. Only EBVs with reliability (R^2) greater than 55% and 20% (without considering the reliability of the breed effect) were used from Models 1

and 2 respectively, where reliability was defined as the square of the correlation between true and predicted breeding values. Reliability was calculated as 1-PEV/V_a, where PEV was the squared standard error of the BLUP estimate. The lower threshold for reliability of the EBVs from Model 2 was used because of the smaller number of records in the dataset leading to generally lower reliabilities for all horses. For the correlation between models, EBVs from Model 2 were re-calculated by adding back the effect of breed group.

2.4 Results

2.4.1 Summary of data

A summary of the dataset analysed for Model 1 is given in Table 2.2, and the distribution of the percentage score for Model 1 is given in Figure 2.1. For Model 1, a total of 6622 competing horses were included in this dataset, of which 6247 were geldings and 375 stallions. Genetic connectedness, primarily due to the sire part of the pedigree was sufficient for parameter estimation. There were 3802 sires, with a mean of 1.7 progeny per sire, 196 sires with >4 progeny and 162 sires with their own competition results within the data. By comparison, connectedness due to the dam pedigree information was less (5303 dams, mean 1.1 progeny per dam, 1 dam with >4 progeny and 4 dams also competing).

Table 2.2 The number of competition records in total, per horse and per event and the number/ mean, standard deviation (s,d,), minimum and maximum (where appropriate) of model variables for Models 1 and 2.

	Model 1				Model 2			
	No/ *mean*	**s.d.**	**Min**	**Max**	**No/ *mean***	**s.d.**	**Min**	**Max**
Total competition records	144509	-	-	-	51391	-	-	-
Horses	6622	-	-	-	2183	-	-	-
Events	16259	-	-	-	13379	-	-	-
Classes	728	-	-	-	705	-	-	-
Competition records per horse	*21.82*	25.87	1	224	*23.54*	26.35	1	154
Competition records per event	*8.89*	6.13	1	65	*3.84*	2.86	1	32
Age (yrs) at competition	*9.63*	3.23	4	30	*8.93*	2.90	4	25
Height (cm)	*165.56*	6.70	121.9	190.5	*165.78*	6.79	121.9	185.4
Percentage (performance trait)	*63.61*	3.36	-	-	*63.84*	3.40	-	-

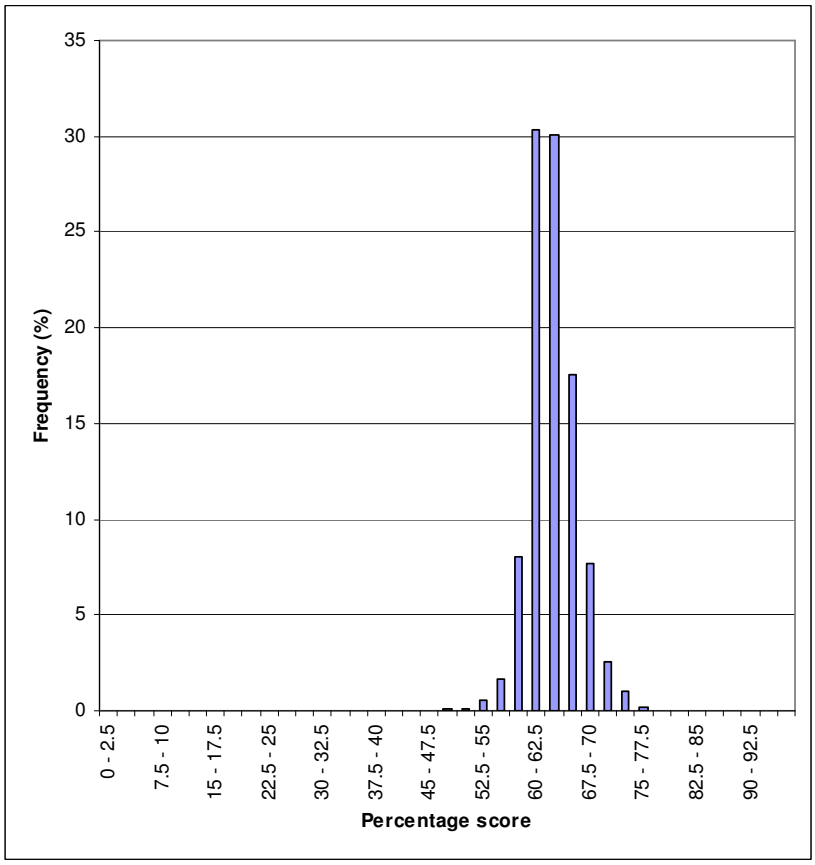


Figure 2.1 Distribution of the percentage score for Model 1. The total number of records was 144509

2.4.2 Genetic effects on dressage performance

Genetic information is obtained from the estimates of V_a and the breed effects. Table 2.3 shows the estimates of heritability. For Model 1, in which the breed classification was ignored, the estimate of h^2 was 0.152 (standard error 0.0178), a result that is statistically significantly different from zero ($P<0.001$). When breed type was explicitly fitted in Model 2, the estimate of h^2 was reduced to 0.110 (standard error 0.036). Comparison between these two estimates is difficult since the second was obtained from a subset of the data used to obtain the first estimate and so the values are not independent. However the results of Model 2 showed breed to be a significant source of variance (see Table 2.4) and therefore the decrease in heritability observed between the two models is to be expected. This was confirmed by the non-breed model restricted to the breed data subset, where heritability increased from 0.110 to 0.144.

The breed comparison showed that the Warmblood/ sport horses performed better than all other breed groups, and this difference was statistically significant for all comparisons other than with the Arab breed. However the small number of Arab horses made this comparison very imprecise. Among the other breed groups the Non-British native/indigenous horses had the lowest performance estimate.

To consider whether the inclusion of the fixed effect of horse height in the model was removing genetic variance, Model 1 was repeated, excluding the fixed effect of height. The inclusion of height had no effect on the estimates of variance components (Table 2.5).

Table 2.3 Estimates of fractions of phenotypic variance explained by each variance component, for Models 1 and 2.

	Model 1		Model 2	
Variance Source	**Ratio to V_p**	**s.e.**	**Ratio to V_p**	**s.e.**
Horse – genetic (V_g)	0.152	0.0178	0.110	0.0364
Horse – permanent environment (V_{pe})	0.115	0.0170	0.172	0.0360
Class	0.025	0.0020	0.027	0.0026
Event	0.055	0.0018	0.066	0.0034
Class.Event interaction	0.100	0.0029	0.076	0.0065
Residual	0.553	0.0046	0.548	0.0090
Total phenotypic variance (V_p)	*12.025*		*12.250*	

Heritability is the ratio of genetic variance to phenotypic variance.
For the non-breed model there were 78423 levels within the class-event interaction, giving an average of 1.84 records per interaction level. For the breed model there were 39368 levels of interaction, i.e. 1.31 records per level.

Table 2.4 Effects of horse breed and competition standard on percentage awarded.

	Model 1			Model 2		
Effect & level	**Est.**	**s.e.d.**	***P*-value**	**Est.**	**s.e.d.**	***P*-value**
Breed						
Warmblood/ sport horse	-	-	-	0.000	-	-
British native/ indigenous	-	-	-	-0.594	0.202	<0.05
Irish	-	-	-	-0.845	0.145	<0.001
Non-British native indigenous	-	-	-	-2.233	0.376	<0.001
Thoroughbred	-	-	-	-0.986	0.155	<0.001
Arab	-	-	-	-0.310	0.467	NS
Standard						
Novice	0.000	-	-	0.000	-	-
Elementary	-1.080	0.067	<0.001	-1.074	0.080	<0.001
Medium	-2.053	0.075	<0.001	-2.108	0.095	<0.001
Advanced Medium	-2.596	0.102	<0.001	-2.672	0.136	<0.001
Advanced	-3.580	0.122	<0.001	-3.533	0.170	<0.001

The control classes for Breed and Standard were Warmblood/ sport horse and Novice respectively. Standard errors shown are the standard errors of the difference (s.e.d.) between each class and the control class.

Table 2.5 Estimates of fractions of phenotypic variance explained by each variance component, for Model 1, excluding the fixed effect of height.

	Model 1, excluding height	
Variance Source	**Ratio to V_p**	**s.e.**
Horse – genetic (V_g)	0.154	0.0178
Horse – permanent environment (V_{pe})	0.115	0.0170
Class	0.025	0.0020
Event	0.054	0.0018
Class.Event interaction	0.100	0.0028
Residual	0.552	0.0046
Total phenotypic variance (V_p)	*12.056*	

Heritability is the ratio of genetic variance to phenotypic variance.

2.4.3 Sources of variation in dressage performance other than genetics

Table 2.3 gives the estimates of the variance components considered in Models 1 and 2, all of which were statistically significant from zero. The effect of fitting breed had most impact on the estimate of permanent environmental variance which increased from Model 1 to Model 2, but as explained above in relation to the genetic variance, a formal statistical comparison of this change is not straightforward. The permanent environmental variance component decreased from 0.172 to 0.147 of the total variance when Model 1 was run using the reduced dataset for Model 2. Among the random terms considered, those attributable to the horse, i.e. the genetic and permanent environment, explain more variance than the combined effects of events, classes within standards and their interaction. The repeatability, which measures the combined effect of the genetics and the permanent environment was estimated as 0.267 (standard error 0.0048) and 0.282 (0.0083) for Models 1 and 2 respectively; note the latter estimate excludes the effect of breed group due to the fact that breed group was accounted for as a fixed effect in the model. Therefore the qualities of the horse account for more than ¼ of the variance observed with approximately 55% of the variance remaining unexplained in the residual.

2.4.4 Effects of gender and age of horse, and competition standard

The age of the horse was also significantly associated with performance ($P<0.001$) as shown in Figure 2.2. For both models, performance was predicted to peak at age 10 then decrease until age 20. The fitted values show an increase among the oldest horses, although the precision of these estimates is low. If this increase were confirmed, a possible reason could be selection; as it is unusual for horses to compete at these advanced ages and only the most consistent horses will continue to compete. Percentage estimates for stallions were significantly higher than those for geldings for Models 1 and 2 ($P<0.001$) and the estimated difference was 1.057 (standard error 0.106) and 1.372 (standard error 0.178) respectively. The better performance by stallions is expected, given that these are an elite selection of males that have been kept for breeding due to good conformation, temperament and competition performance. There was a clear trend in decreasing percentage awarded with increasing standard of the competition test ($P<0.001$, see Table 2.4), so that highest scores tended to be observed in the Novice standard. This is possibly due to less strict judging at the lower ability standards.

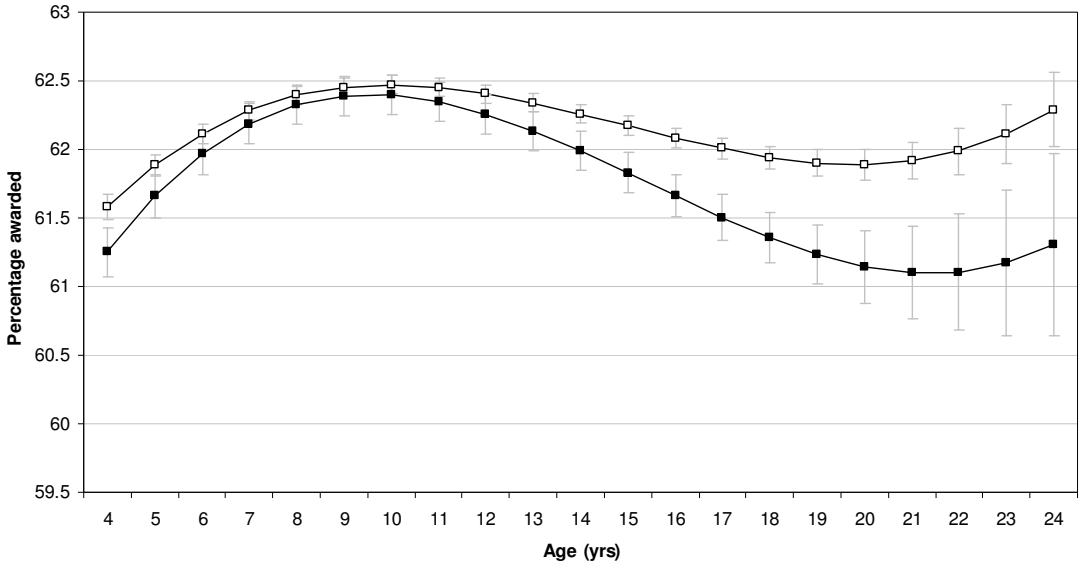


Figure 2.2. Predicted percentage at different ages from Model 1 (open squares) and Model 2 (closed squares) (+/- 1 standard error bars). Where all other fixed effects are averaged. Value at age = 24 is for age group 24 – 30 years. The fitted means of the curves are slightly different due to differences between the fixed effects models.

2.4.5 Effect of the height of horse

In Model 1, height was significantly associated with performance, showing a curvilinear relationship ($P<0.05$), as shown in Figure 2.3. The relationship showed a sickle-shaped concave curve. Above approximately 142 cm, percentage increases with increasing height, in an almost linear fashion, whereas below 140 cm there was little evidence of a relationship. At heights below 142 cm, the large error bars indicate a lack of precision due to scarcity of data. However, much of this relationship appears due to the different breeds, as removing the variance due to breed in Model 2, height became non-significant. A more detailed analysis fitting separate curves for each breed is shown in Figure 2.4, but note that despite the range of relationships with

height exhibited by the different breeds, these differences were not statistically significant, with the exception of the Warmbloods for which there was a significant linear relationship suggesting improved performance with increasing height. This indicates that height effects within breed are small, and that the apparent effect of horse height on performance may be due to confounding with effects of breed, in that Warmbloods, which typically perform better at dressage as shown in Table 2.4, primarily appear towards the upper end of the range of heights for competing horses.

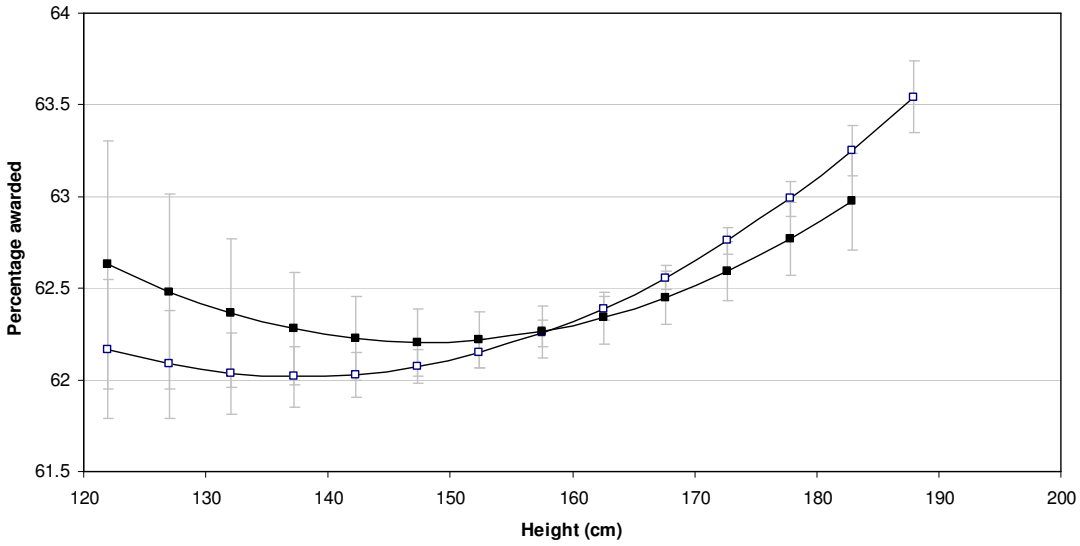


Figure 2.3. Predicted percentage by height from Model 1 (open squares) and Model 2 (closed squares) (+/- 1 standard error bars). Where all other fixed effects are averaged

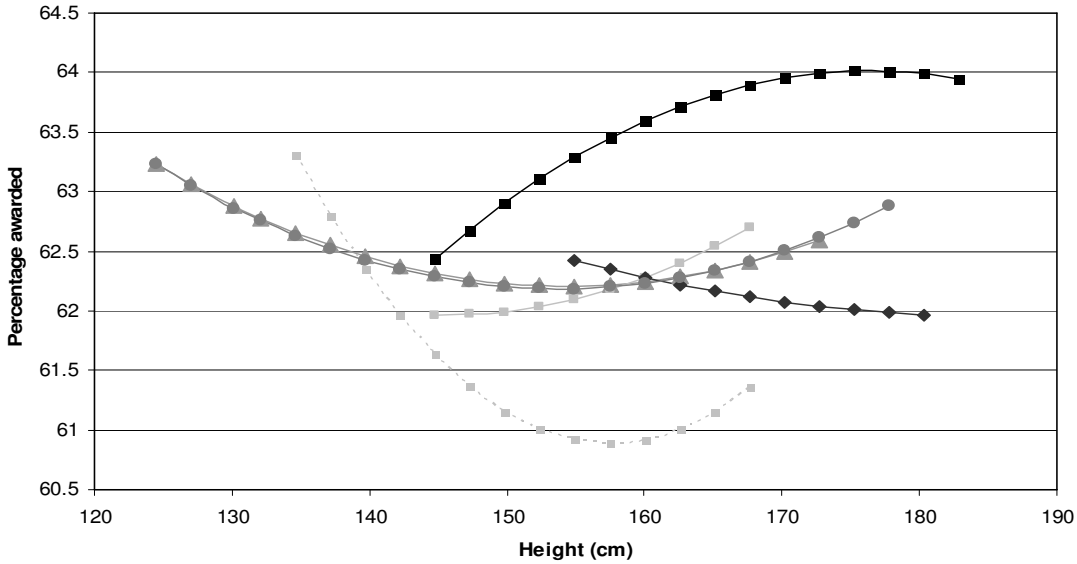


Figure 2.4. Predicted percentage by height for each breed in Model 2, plotted over 90% height range for breed. Curves are shaded with grading according to the number of records in each group. The black represents Warmbloods and lightening shades of grey represent, Thoroughbreds, Irish, British native/ indigenous, Non-British native/ indigenous and Arabs respectively.

2.4.6 Estimated Breeding Values

The correlation between estimated breeding values derived from the two models was 0.948, based on 450 horses, and a reliability of >55% and >20% (without considering the reliability of the breed effect) in Models 1 and 2 respectively (Figure 2.5). For EBVs calculated from Model 1, the distribution of EBVs for the 113 sires with a reliability of >55% and the genetic trend over time for all competing horses with an EBV reliability of >55% is shown in Figures 2.6 and 2.7 respectively. This distribution was slightly right skewed, but otherwise had a typical bell-shaped distribution. Note that with no minimum reliability imposed, and including EBVs for all horses in the pedigree, the distribution was bell-shaped and centered around zero; a bigger proportion of better horses (on average) was represented by sires with high EBV reliability. Sires with greater reliability tended to have higher EBVs For the trend in EBVs over time, the mean EBV for all

competing horses have high standard errors in the extremes of the time range, since there is less data available for the oldest and youngest horses at present. Nevertheless, there appears to be an increase in EBV over the period covered, with an increase most evident from around 1990. A simple (weighted) regression of individual EBV on year of birth gives an estimate of genetic progress between 1985 and 2001 of 0.047 (standard error 0.021) genetic standard deviations per annum. The trend in phenotype over time was very similar to the genetic trend (results not shown).

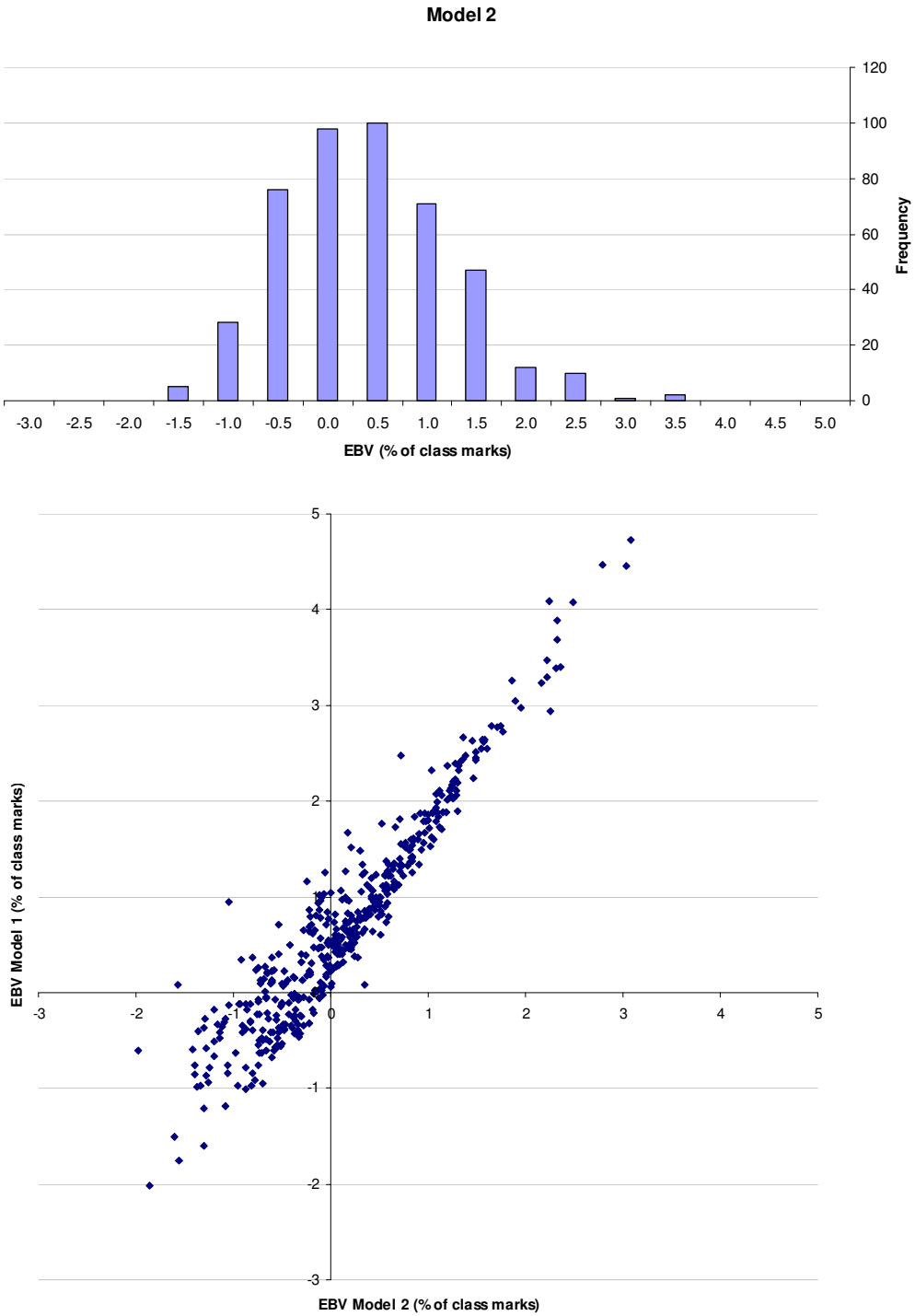


Figure 2.5 Correlation between EBVs in Model 1 and Model 2 (including breed effect) with histogram of the frequency of horses by EBV in Model 2. Reliabilities of EBVs were >55% and >20% (without considering the reliability of the breed effect) for models 1 and 2 respectively.

2.5 Discussion

Breed significantly affects dressage competition performance, and ideally, EBVs within the GB population will take this into account. This analysis accounted for breed by grouping into (six) robust categories that would be of use for GB breeders, after consultation with British Breeding. The model including breed is the most appropriate, giving a more informative estimate, having removed variance due to differences between the breed groups. However relative to Model 1, the breed model was based on approximately a third of the data, and the reliability of EBVs from the breed model was particularly low. The correlation of EBVs between the two models was 0.948, indicating that addition of breed changes the EBV estimates and rankings of horses, but only slightly. This was based on a reliability of >55% and >20% (without considering the reliability of the breed effect) in Models 1 and 2 respectively; if this restriction was removed it would lower the correlation. Given the current limitations with the breed data, it may be possible to use Model 1 as an interim measure in practice without resulting in a large loss of progress. Nevertheless, it would be straightforward to capture an extra 25% of the genetic variation by the routine recording of breed (according to predefined criteria, such as the breed classes used here).

Few routine genetic evaluation models account for differences between breeds (Thorén Hellsten *et al*., 2009b; Janssens, 2008), although some models currently take percentage Thoroughbred into account (Janssens, 2008). In this study, the Warmblood/ sport horse performed better than all other classes, with the exception of the Arab, where no significant difference was detected. Given the intended breeding goal of the sport horses, with suitable temperament and athletic abilities, this is not surprising. The Irish category consisted of those registered with the Irish studbooks (excluding thoroughbred where known), so may have included thoroughbreds, Irish Drafts, Irish Sport Horses and others. Irish Sport Horses have been bred to compete particularly in showjumping, rather than dressage, which may explain the lower

performance of this group. Thoroughbreds are bred for speed, while British native/indigenous horses such as Highland ponies and the Cleveland Bay are bred for stamina/strength and as draft horses respectively. Both these classes had significantly lower performance estimates. The Non-British native/indigenous group included a range of breeds, that either are not bred to compete in dressage (e.g. the American Quarter Horse), or are bred to compete, but may attract negative judging bias (e.g. the Andalusian), as a group therefore, the low performance is comprehensible. In a study comparing heritabilities and genetic correlations of traits recorded at mare tests for a number of German Warmblood breeds, Lüehrs-Behnke *et al*., (2006b) found large differences in the additive genetic variance of the traits between breeds. Heritability estimates from this analysis are of the same magnitude as estimates from evaluations performed internationally. Compared to international studies (which generally relate to one studbook) the heritability estimate from the model including breed is consistent with other heritabilities that have been published for adult dressage competition performance. For example, heritabilities ranging from 0.10 to 0.20 have been estimated in various breeds (see Janssens, 2008; Olsson *et al*., 2008; Ducro *et al*., 2007; Ricard *et al*., 2000), although there is little indication of the precision of many of these estimates. For many of these estimates, the competition performance measure used relates to performance in one event, rather than an aggregate or cumulative measure over competitions or years for example those used in the KWPN and SF evaluations. A mean measure based on a number of competitions would predictably produce an increased heritability reflecting the decreased residual variance.

The evaluation based on the single competition however, has increased information and is more flexible as a summary statistic. It allows for variation in the number of competitive records per horse and can allow for event related variables e.g. course, and date related measures such as weather, and rider if available (Kearsley *et al*., 2008). For our data, assuming that a horse

competes in 4.6 competitions per year (the observed average number of competition records per horse, per year, with a career of 5 years) the heritability will increase to 0.27 (for Model 1). This annual estimate is more comparable to the estimate of 0.34 obtained in the SF population using the log of annual earnings (Janssens, 2008). Kearsley *et al*. (2008) estimated 0.09–0.11 for the dressage phase of GB eventing competitions, without accounting for breed differences. Comparing the non-breed model estimate (0.15) to the Kearsley *et al*. (2008) estimate, our estimate is higher. This may be due to the fact that we were unable to remove the rider variance, in contrast to Kearsley *et al*. (2008), where rider contributed up to 18% of the phenotypic variance. However, in the present analysis, the genetic variance primarily comes from differences between sires and their half-sib groups, therefore rider variance is unlikely to have contributed to the heritability. When breed was included as a random rather than a fixed effect in the model (results not shown), the estimate of variance due to breed is approximately 4% of the total phenotypic variance, corresponding with the decline in heritability between Models 1 and 2.

This analysis used competition data and so brings with it the problem that it is likely to have been influenced by selection. There is likely to be selection in the competing population — some horses may never compete, or may have gaps in their careers due to breeding purposes. The degree of selection will increase at higher ages and levels of competition. However, a relatively large proportion of the population should have been included, as many horses will compete at some point. The system of recording results has introduced censoring in that values below 60% are not routinely recorded. Therefore in addition to selection of horses that compete, there is selection of performances for an individual horse, where the worst performances are not taken into account when estimating the genetic value. These limitations are general problems in the use of competition data for evaluations, and inherent in almost all evaluations internationally.

Further selection occurred in our evaluation as pedigree records were available for only a selection of horses with records. The consequence of the censoring in data recording will be to decrease heritability estimates, with the degree of underestimation determined by the degree of truncation. Changes in data recording – by recording all results as a matter of routine – would therefore increase heritability estimates and lead to more accurate estimates of breeding values. There were further selective practices in the recording of mares, but our estimates avoid this bias due to their exclusion from the analysis.

The inclusion of height in Model 1 may act primarily as a proxy for breed; height but not breed data was available for the whole dataset, and height has been shown to be heritable for many breeds and dependent on the breed. Heritability estimates include 0.59 for the Swedish Warmblood, 0.79 for the Haflinger, 0.25 in the Trakener, 0.48 in the Arab (Gerber *et al*., 1997; Miglior *et al*., 1998; Kaiser *et al*., 1991; Seidlitz *et al*., 1991, all as cited by Saastamoinen & Barrey, 2000), 0.89 for the Shetland pony (van Bergen & van Arendonk, 1993), 0.47 for the Connemara pony and 0.34 for the New-Forest pony (Ricard, 2004). Studies also indicate that height is positively genetically correlated (0.33 and 0.50) with performance in dressage (Ducro *et a*l., 2009; Viklund *et al*., 2008). Therefore, if height was included as a routine fixed effect in the model (without acting as a proxy for breed) it suggests that an underestimate of heritability may result. In the absence of well recorded performance data, the information from height could be assimilated via a bivariate analysis. Analyses were repeated having removed height as a fixed effect, however there was no discernible effect on the heritability and variance component estimates.

There are many other factors that influence performance of a horse and these are accounted for by the other factors in the model. Our model also accounted for the permanent environment of

the horse, which contributed 11.5% of the phenotypic variance. This represents the effect of the stables, upbringing, training, nutrition, maternal effects etc. It is likely that much of the rider variance will be accounted for in this term as there is a tendency for a rider to compete with a particular horse. Competition class accounted for 2.5% of the total variance and the event 5.5%. The event represents effects of temporary environment—e.g. course conditions, course surface, arena and weather. The event–class interaction accounted for a relatively large 10.0% of the variance, suggesting there are substantial differences between events for the same class at different events, after accounting for the effect of the event, possibly due to factors such as judging differing between events. Results for the effect of age are consistent with other studies, which generally find that performance increases with age, or levels off (Kearsley *et al*., 2008; Koenen & Aldridge, 2002). To examine this further, separate curves of predicted percentage were fitted for each standard (results not presented). These indicated that age of peak performance increased with advancing standard of competition; for Novice and Elementary performance peak age was 9 years, for Medium it was 10 and for Advanced Medium and Advanced performance peaked at age 12. Other studies generally show that stallions and geldings perform better than mares (Koenen & Aldridge, 2002).

Genetic progress was examined considering horses with a reliability of >55%, which included 17% of the competing horses, and 3% of sires. Ideally, a higher reliability would be used, if this were available for a large enough proportion of the evaluated horses. However, the reliability of EBVs estimated in this analysis is fairly low, due to the limited extent of pedigree and number of competition records. The increase in average EBVs between 1985 and 2001 is 0.046 genetic standard deviations per year, which is consistent with estimates for jumping and gaits in the Swedish Warmblood population, between 1988 and 1992 (Olsson *et al*., 2000). This indicates that genetic progress has been made in the UK sport horse population for dressage competition,

in the past 15+ years, due to the current breeding practices. This may be due to a greater importation of horses from the continent, where the introduction of BLUP has promoted genetic progress in these populations. Evidence for the impact of BLUP is also seen in the Swedish Standardbred trotter and Icelandic horses (Árnason, 1997; Sigurdsson *et al*., 1997, both as cited by Árnason & Van Vleck, 2000). Evidence for significant genetic progress, assessed by temporal trends in cohort mean EBVs and estimated using regression of cohort mean EBV on year, must be treated with some caution, as illustrated in an evolutionary context (Hadfield *et al*., 2010). Significance tests are very anti-conservative due to autocorrelation between PEVs between relatives, and thus positive correlations between successive cohort mean BLUPs, and also due to drift resulting in positive correlations between cohort mean BVs. The use of the breeding value complete posterior distribution determined in a Bayesian analysis was advocated as an alternative way to assess temporal changes. In the current context, however, the consequences of these issues may be minimal. The dataset included a wide range of, and relatively long generation intervals (e.g. 4 to 12 years) over a a time period spanning a few generations, thus correlation between successive cohort mean EBVs will be less influenced, and drift in the population over these few generations is also likely to be minimal.

The introduction of EBVs for the UK sport horse population will further augment the current breeding initiatives, to enable greater genetic progress and provide every opportunity for the UK to be internationally competitive.

2.6 Conclusion

This is the first study to estimate genetic parameters and perform genetic evaluations for performance in dressage in GB. The heritability estimates of 0.15 and 0.11, without and with breed groups in the model, indicate that it should be possible to predict BLUPs of sufficient

reliability and perform selection based on these values in the GB population, a practice which is not currently conducted.

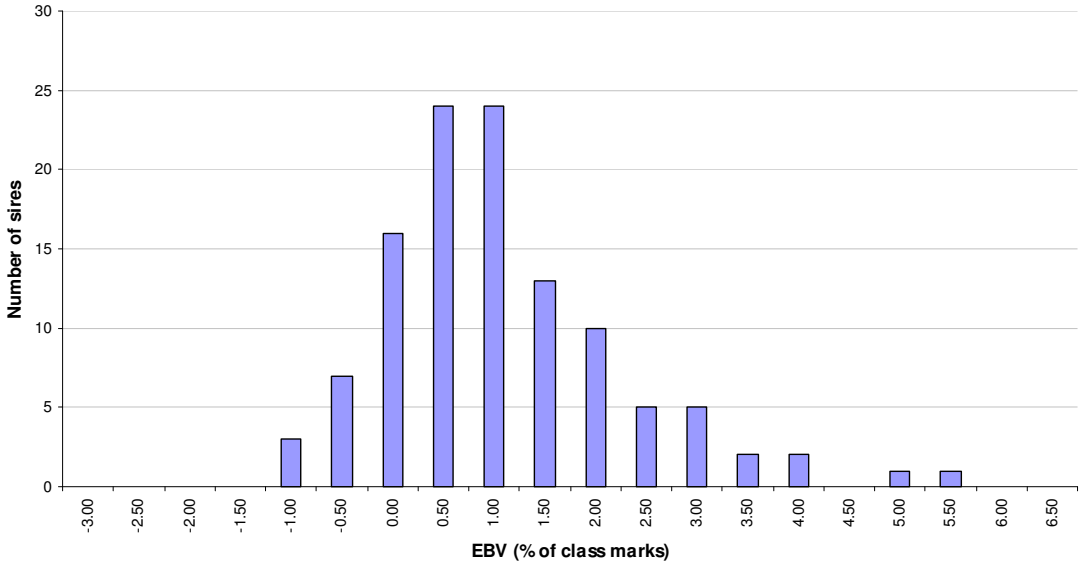


Figure 2.6. Distribution of sire EBVs, from Model 1. The EBV interval labelled 1.00 represents EBVs from >0.50 to 1.00 and correspondingly for the other labels.
Reliabilities for sire EBVs were >55%.

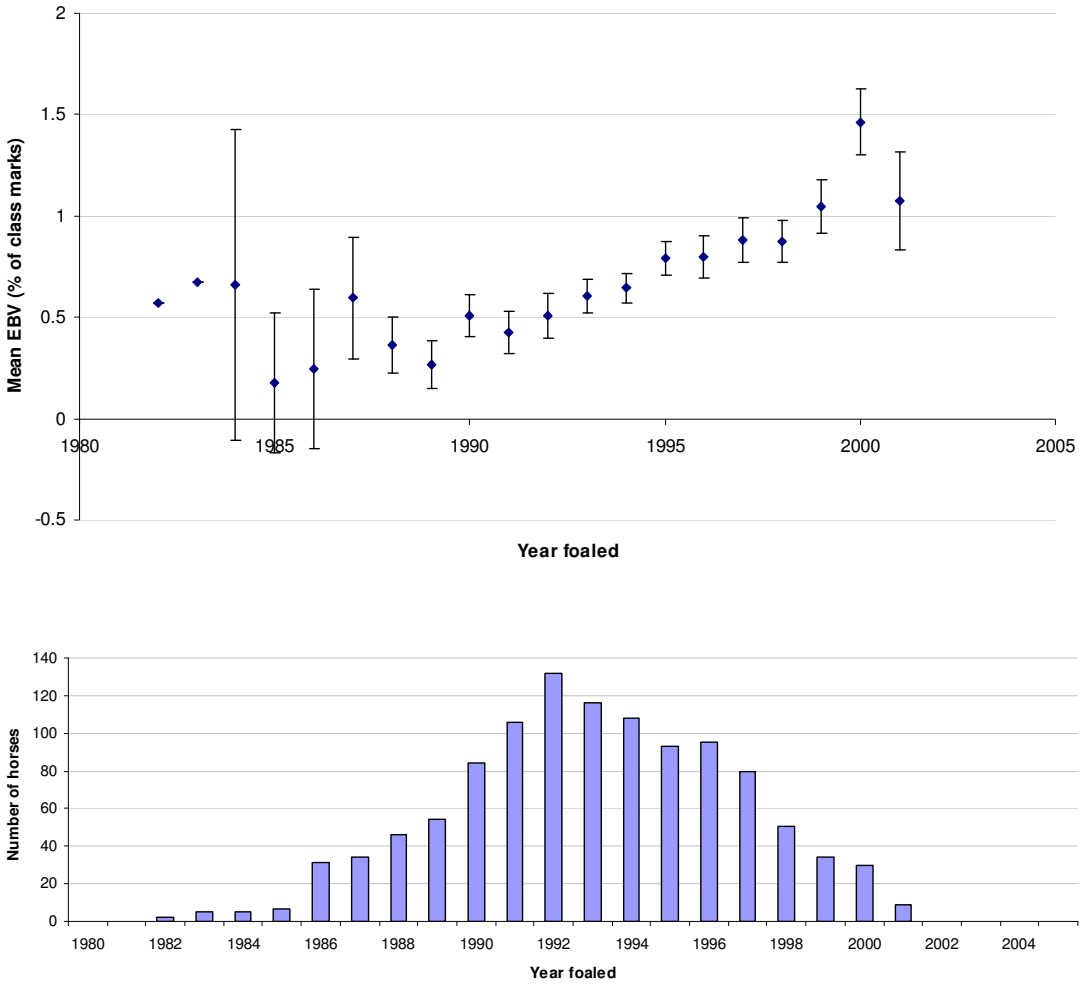


Figure 2.7a Trend in mean EBV, from Model 1, over time. The mean EBV is for all competing horses with > 55% reliability. Bars represent +/- standard errors. b Number of horses used to derive mean EBVs in Figure 2.7a, according to year foaled.

3. Estimating variance components and predicting breeding values for eventing disciplines and grades in sport horses

3.1 Abstract

Eventing competitions in GB comprise three disciplines, each split into four grades, yielding 12 discipline-grade traits. As there is a demand for tools to estimate (co)variance matrices with a large number of traits, the aim of this work was to investigate different methods to produce large (co)variance matrices using GB eventing data. Data from 1999 to 2008 were used and penalty points were converted to normal scores. A sire model was utilized to estimate fixed effects of gender, age and class, and random effects of sire, horse and rider. Three methods were used to estimate (co)variance matrices. Method 1 used a method based on Gibbs sampling and data augmentation and imputation. Methods 2a and 2b combined sub-matrices from bivariate analyses; one took samples from a multivariate Normal distribution defined by the covariance matrix from each bivariate analysis, then analysed these data in a 12-trait multivariate analysis; the other replaced negative eigenvalues in the matrix with positive values to obtain a positive definite (co)variance matrix. A formal comparison of models could not be conducted; however estimates from all methods, particularly Methods 2a/b were in reasonable agreement. The computational requirements of Method 1 were much less compared to Methods 2a or 2b. Method 2a heritability estimates were, for dressage 7.2 to 9.0%, for show jumping 8.9 to 16.2%, and for cross-country 1.3 to 1.4%. Method 1 heritability estimates were higher for the advanced grades, particularly for dressage (17.1%) and show jumping (22.6%). Irrespective of the model, genetic correlations between grades, for dressage and show jumping, were positive, high and significant, ranging from 0.59 to 0.99 for Method 2a and 0.78 to 0.95 for Method 1. For cross-

country, by Method 2a, genetic correlations were only significant between novice and pre-novice (0.75), however, by Method 1 estimates were all significant and low to moderate (0.36 to 0.70). Between discipline correlations were all low and of mixed sign. All methods produced positive definite 12*12 (co)variance matrices, suitable for the prediction of breeding values. Method 1 benefits from much reduced computational requirements, and by performing a true multivariate analysis.

3.2 Introduction

Eventing is the equestrian sport in which the horse and rider compete in each of the three individual competition disciplines – show jumping, dressage and cross-country. Internationally, genetic evaluations for performance in sport horses tend to focus on the individual disciplines of show jumping and dressage, for which many studbooks produce annual breeding value estimates. Genetic evaluations for eventing competition are rare. Langlois (1980) and most recently Ricard & Chanu (2001) performed evaluations in the French population, the latter using earnings and ranks of the overall competition, and Kearsley *et al.* (2008) performed an evaluation in the GB horse population, using penalty points, and analyzing individual disciplines as well as an overall trait. The heritability of overall competition in the French population was 0.11 / 0.17 (annual results; natural log of earnings per number of starts and per number of places respectively) and 0.07 (rank in each event) (Ricard & Chanu, 2001). The heritability in the GB population was 0.09–0.11 for dressage, 0.08–0.23 for show jumping, 0.02–0.03 for cross country and 0.05 for overall competition, using the performance measure of penalty points in a single competition (Kearsley *et al.,* 2008). These estimates for the show jumping and dressage phases were similar to those derived from evaluations of the individual disciplines.

In GB eventing competition data, there are 12 traits for the combinations of discipline (3) and grade (4; which are, in order of increasing ability, pre-novice, novice, intermediate and advanced). Overall performance in a competition is based on a sum of the scores for the individual disciplines. Ideally genetic evaluations would be based on a 12 trait model. To enable this, genetic parameters for each of the discipline-grades, and correlations between them would need to be estimated. Separate estimated breeding values for each of the 12 competition-grades could then be predicted for each horse. These would be available for all traits, or could be combined into an aggregate index, for example, for each discipline, or for competition overall. The production of 12 traits, or an index based on these, is of interest as it allows flexibility depending on the breeding goals. For example, breeders may have a particular interest in breeding for a specific level of competition, i.e. advanced for the professional rider, or novice for the amateur rider, or the interest may be in producing a horse for competing in one or more of the disciplines.

Genetic evaluations for sport horses typically estimate variance components using residual maximum likelihood (REML), and then predict breeding values using best linear unbiased prediction (BLUP). Most BLUP programs require estimated (co)variance matrices which are positive definite. When the whole matrix is being estimated directly, this frequently results in non-positive definite matrices (negative eigenvalues), and the probability of this increases as the number of traits increases (Hill and Thompson 1978). This can be avoided by imposing constraints to keep the matrix positive definite. However, to cope with the computational demands of estimating (co)variance matrices for a large number of traits, often (co)variance matrices are calculated for subsets of the traits, and the estimates are then combined. Again, this frequently results in non-positive definite matrices and particular methods are required to combine the submatrices and produce a positive definite full matrix. Such methods include

Mantysaari (2004), which uses a random regression model modified to use a rank deficient sire (co)variance matrix, Wall *et al.* (2005) which introduced a method based on the Cholesky decomposition, minimising the Frobenius distance (Higham 2002; Sorensen *et al.* 2002), among others (Knol and Ten Berge 1989). A disadvantage of this approach is that combining submatrices to form a full (co)variance matrix does not make best use of the available information. Alternatives to REML include Bayesian methods, such as Markov Chain Monte Carlo simulation (MCMC), which has been used in horse populations (Stock *et al.* 2007). A comparison of Bayesian methods and REML is discussed in detail in Misztal (2008). A feature of MCMC is that it is suitable for multivariate analysis, however may be slow to compute (Misztal, 2008).

This study was driven by the need to develop comprehensive genetic evaluations of horses competing in eventing competitions in GB using the information from all 12 traits in a multivariate fashion. However the problems encountered in obtaining a positive definite 12x12 (co)variance matrix prompted the study to include a comparison of different methodologies.

3.3 Materials and Methods

3.3.1 Data

Competition results, from 1999 to early 2008, were obtained from British Eventing, the body regulating the sport in GB. Results from GB competitors competing in national and international competitions were included. The performance traits used were the penalty point scores for each of the three individual disciplines. Penalty points were converted to normal scores within the competition class, using the method of Kearsley *et al.* (2008), adapted from Royston (1982). The better performing horses were awarded fewer penalty points, and so a negative Normal score represented a better performance than a positive score. The competition class grouped

competitors competing against each other at the same event and covers the standard of the competition, date, location and related temporary environmental factors such as weather. The transformation of penalty points to Normal scores achieved a distribution closer to the Normal, with a zero mean score for each competition class. Horses less than 4 years old were excluded, and horses 20 years or older were grouped into a single age group.

The ability-grade combinations were as follows: dressage pre-novice (DP), dressage novice (DN), dressage intermediate (DI), dressage advanced (DA), show jumping pre-novice (SJP), show jumping novice (SJN), show jumping intermediate (SJI), show jumping advanced (SJA), cross-country pre-novice (XCP), cross-country novice (XCN), cross-country intermediate (XCI), and cross-country advanced (XCA). In the following discussion, those 12 combinations were the 12 traits.

The dressage phase is always performed first, followed by either show jumping or cross-country, depending on the nature of the event. Therefore, the dressage data represented the full number of competition records; subsequent exclusions during the latter phases resulted in slightly less records for them. Records where penalty points did not follow this pattern of performance related exclusion were removed.

Horses and riders often appear in multiple grades in the dataset as a whole. Conversely, some horses may not have competed in all (i.e. the latter) phases of competition due to exclusions, or within every grade, however, due to genetic connections between animals (in this case limited to connections within half-sib groups) and genetic associations between grades and disciplines, breeding values can be predicted for all horses at all grades and disciplines. The rider was considered an important source of variation (Kearsley *et al.,* 2008) and is estimable because

riders are associated with more than one horse (mean number of horses per rider = 2.8), and horses are also commonly associated with more than one rider (mean number of riders per horse = 1.7). Rider has been recorded consistently since 1999; data before this time where rider was intermittently recorded was excluded. The sire of competing horses was available from British Eventing, and was generally well recorded. Sires were recorded by name however, with no other unique identifiers. Data was cleaned manually, conservatively grouping sire names where they were considered to be the same animal. Variance components were estimated using a dataset reduced in size to include sires with either i) 30 or more individual progeny competing in either pre-novice or novice grades or ii) any progeny in intermediate or advanced grades.

3.3.2 Models

The basic model was a sire model with mixed linear effects, fitted within each discipline-grade:

$$y = \text{mean} + \text{gender} + \text{age} + \text{age}^2 + \text{class} + \text{sire} + \text{rider} + \text{horse} + e$$

The fixed effects were gender of the horse ('gender': stallion, geldings and mares), the age of the horse at time of competition fitted as linear and quadratic covariates (based on a preliminary analyses) and competition class ('class'). Random effects were the additive genetic effect of the sire ('sire'), the rider ('rider'), the permanent environment of the horse ('horse') and the residual error (e). The random effects of the sire, rider, horse and residual error were assumed normally distributed with (co)variance matrices $\Sigma_{sire} \otimes I$, $\Sigma_{rider} \otimes I$, $\Sigma_{horse} \otimes I$ and $\Sigma_e \otimes I$.

Due to computational constraints it was impossible to perform a standard 12-trait multivariate analysis for the full model, so three alternative methods were used to estimate the 12x12

(co)variance matrices. The relationship between traits was assumed to be unstructured, i.e. there were no specified correlations between grades.

Method 1: A method called data augmentation was used to perform a 12-trait multivariate analysis. This data augmentation is based on work by Thompson (1994) and Clayton and Rasbash (1999) where computational requirements are reduced, using data with hierarchically nested random effects, by repeatedly fitting sub-models in an overlapping series, with each sub-model being fitted in turn to data adjusted for effects not in the current sub-model. In an iteration of the full model, an internal iteration of each sub-model is performed. This greatly reduces the computational requirements. Fixed and random effects are updated as they are re-estimated. A simplified form of Gibbs sampling is used to add noise to the updated estimates at each step, thus preventing bias in the estimated effects. For random effects, the noise added to each solution is taken as a sample from a normal distribution with a variance equal to the prediction error variance of that solution; for fixed effects, the noise is sampled from a normal distribution with variance equal to the square of the standard error. The calculations were carried out in a development version of ASReml 3 (Gilmour *et al*., 2009).

For example, if the two sub-models are: 1) $y - Zu = X\beta + e$ and 2) $y - X\beta = Zu + e$, where y is a vector of phenotypic observations, u and β are vectors of random and fixed effects respectively, e is a vector of residual errors, Z and X are design matrices allocating observations to random and fixed effects, the process is as follows:

i) For model 1, assume u is 0, estimate and add noise to β

ii) For model 2, augment the data by subtracting $X\beta$ using an imputed value of β from i), estimate u and variances and add noise to u

iii) For model 1, augment the data y by subtracting Zu using an imputed value of u from estimate of u from ii), estimate and add noise to β

iv) repeat ii and iii, for total number of iterations

v) exclude burn-in iterations, and calculate averages of estimates

In this analysis, the full model was split into 4 sub models. These were: 1) the fixed effect of class within discipline and grade 2) the random effect of sire within discipline and grade 3) the random effect of rider within discipline and grade 4) the random effect of horse within discipline and grade. The fixed effects of gender within discipline and grade and age (linear) and age (quadratic) within discipline andgrade were included in all sub-models. A burn in period of 60 iterations was allowed, and 500 iterations were executed. Residual error covariances were fixed at zero since, as a simplification, it was assumed that there was no covariance between discipline-grades for the remaining un-explained variance. (Co)variance matrices for each random effect were estimated as the average over all iterations excluding the burn-in period, making a total of 440 iterations.

The 12-trait multivariate model (before division into sub-models) was:

$$y = dg.mean + dg.gender + dg.age + dg.age^2 + dg.class + dg.sire + dg.rider + dg.horse + dg.e$$

where y is a matrix of responses and dg is the factor indicating discipline-grade. The fixed effects were gender of the horse ('gender'), age of horse at time of competition (polynomial, linear and quadratic) and competition class ('class'). Random effects were the additive genetic effect of the sire ('sire'), the rider ('rider'), the permanent environment of the horse ('horse') and

the residual error (e). The random effects sire, rider, horse and residual error were assumed normally distributed with (co)variance matrices $\Sigma_{sire} \otimes I$, $\Sigma_{rider} \otimes I$, $\Sigma_{horse} \otimes I$ and $\Sigma_e \otimes I$. The estimated matrices were constrained to be positive-definite by performing an expectation-maximization update if the average information REML update generated a non-positive definite matrix. Estimates for the effect of the discipline-grade were available.

The analysis gives solutions plus their standard errors for all fixed and random effects. Unfortunately, as yet no standard errors are available for estimates of variance components, therefore estimates of standard errors from the bivariate analyses were used (see Methods 2a/2b below). The standard errors from the various models are unlikely to differ much, and we might expect the standard errors from a 12-trait analysis to be lower than the standard errors obtained if only 2 of the 12 traits were included in the analysis. For this model, standard errors of correlations between traits are approximate.

Methods 2a/2b. A series (66) of bivariate analyses, for every possible pair of discipline-grade traits was run using ASReml. Two distinct methods were then used to combine the series of 2*2 submatrices from each strata (sire, horse, rider) into the 12*12 covariance matrices.

Method 2a: For each of the 66 bivariate analyses, 3 pairs of observations were sampled from a multivariate Normal distribution defined by the 2*2 covariance matrix. These were analysed in a 12-trait multivariate analysis using ASReml, where the only fixed effect in the model was a separate mean for each of the 66 bivariate analyses. At least 3 samples were required to ensure that the resulting (co)variance matrix matched the (co)variance matrix from the bivariate analysis. This is an extension of Wall *et al*., 2005, designed to ensure that a mean could be fitted in the final model.

Method 2b: The average (co)variance values for the 66 bivariates were calculated as simple means. To convert these average (co)variance matrices into positive definite matrices, they were decomposed to give the eigenvalues and eigenvectors. Negative eigenvalues were changed to $1*10^{-10}$ and the matrices were re-composed. This method is a simplification of more complex methods which minimize the Frobenius distance between the original estimate and the final positive definite estimate (Higham 2002; Sorensen *et al.* 2002) i.e. equivalent to computing the positive-definite matrix that is best fitting by least-squares (Knol and Ten Berge 1989). Standard errors of correlations were not accounted for in the process. The significance of random effects was examined using an approximate t-test, and where the variance component was obviously significantly greater than zero, no further test was done. A likelihood ratio test was performed for variance components where the significance was less apparent. The significance of fixed effects was examined using an approximate F-test; differences between levels of fixed effects were assessed using the t-test.

Thus, 3 methods were used to produce the positive definite 12x12 (co)variance matrices, suitable for BLUP estimation of breeding values. To estimate the effect of fixed effects for methods 2a/2b, the results from the 66 bivariates were averaged.

3.3.3 Calculation of functions of variance components

Heritabilities (h^2), repeatabilities (r^2), and fractions of variance due to permanent environment (c^2) and rider (t^2) were calculated as functions of the various components using the phenotypic variance defined as $\sigma_p^2 = \sigma_s^2 + \sigma_c^2 + \sigma_t^2 + \sigma_e^2$. The functions were calculated as $h^2 = (4\ \sigma_s^2)/\ \sigma_p^2$, $r^2 = (\sigma_s^2 + \sigma_c^2)/\ \sigma_p^2$, $c^2 = (\sigma_c^2 - 3\ \sigma_s^2)/\ \sigma_p^2$, and $t^2 = \sigma_t^2/\ \sigma_p^2$,where σ_p^2 is the total variance, σ_s^2 is the sire variance, σ_c^2 is the residual horse variance, σ_t^2 is the rider variance and σ_e^2 is the residual variance.

3.4 Results

Table 3.1 summarises the dataset, giving the number of records, horses, sires, number of records by horse gender, and number of riders for each discipline-grade in the dataset. The distribution of penalty point scores (before transformation to Normal scores) are shown in Figures 3.1a – 3.1c, for each of the disciplines individually.

Table 3.1 Number of records, horses, sires, number of records by horse gender and number of riders for each discipline-grade in the dataset. The total number of horses, sires and riders was 19,829, 3,017 and 11,841 respectively.

Discipline-grade	No. records	horses	sires	Number of records by gender			riders
				Stallions	Geldings	Mares	
DA	18413	2202	1308	138	16081	2194	1017
DI	77943	6863	2982	1133	61579	15231	3301
DN	113434	10731	2608	1162	81585	30687	6079
DP	135277	16664	2396	814	90788	43675	10767
SJA	16405	2163	1285	126	14325	1954	989
SJI	66430	6380	2844	934	52843	12653	2969
SJN	116304	10625	2626	1261	83801	31242	5948
SJP	130953	16460	2392	804	87874	42275	10597
XCA	14765	2052	1235	110	12907	1748	945
XCI	59058	6090	2732	836	47052	11170	2810
XCN	106558	10277	2612	1150	76991	28417	5715
XCP	120939	16055	2383	747	81357	38835	10315

DA =dressage advanced, DI=dressage intermediate, DN=dressage novice, DP=dressage prenovice, SJA=show jumping advanced, SJI=show jumping intermediate, SJN=show jumping novice, SJP=show jumping prenovice, XCA=cross-county advanced, XCI=cross-country intermediate, XCN=cross-country novice, XCP=cross-country prenovice

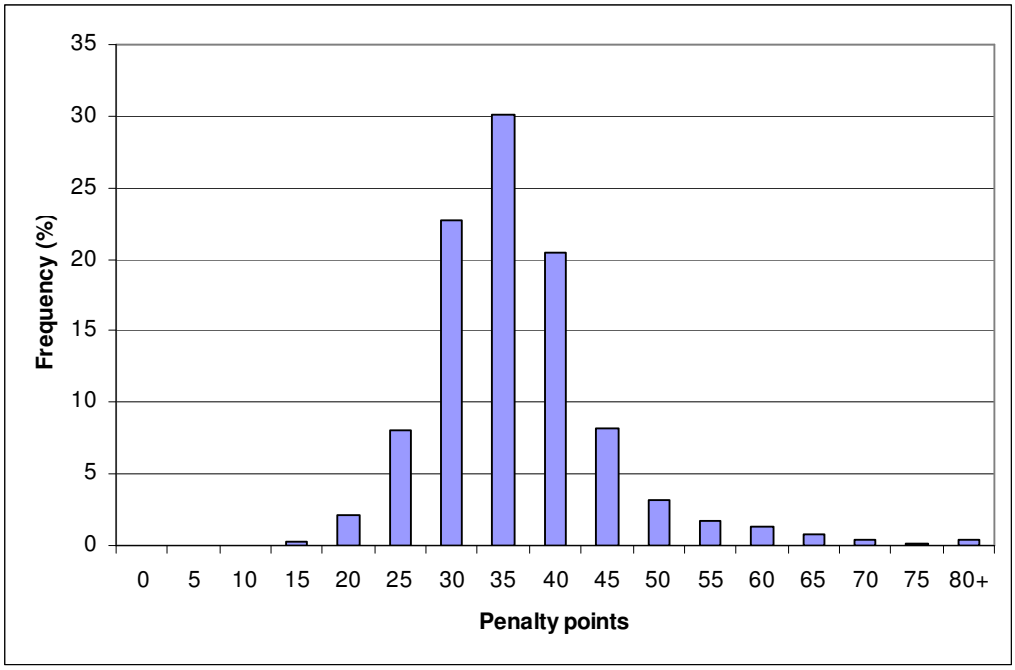


Figure 3.1a Distribution of penalty points for dressage. Total no. records was 345067

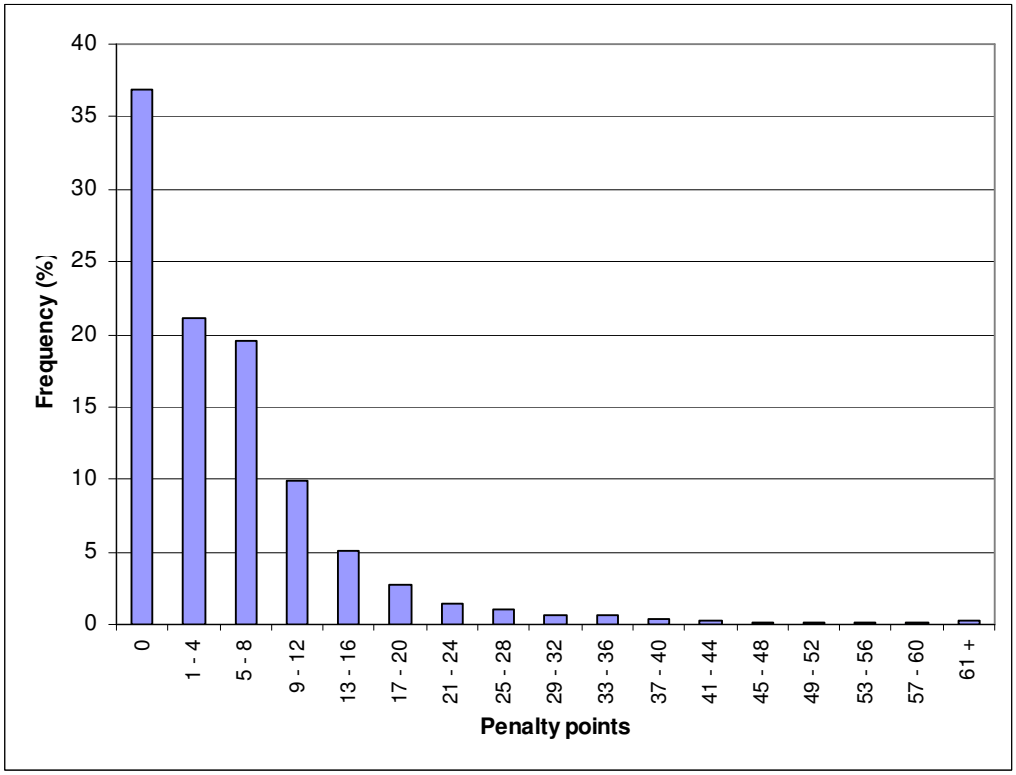


Figure 3.1b Distribution of penalty points for show jumping. Total no. records was 329597

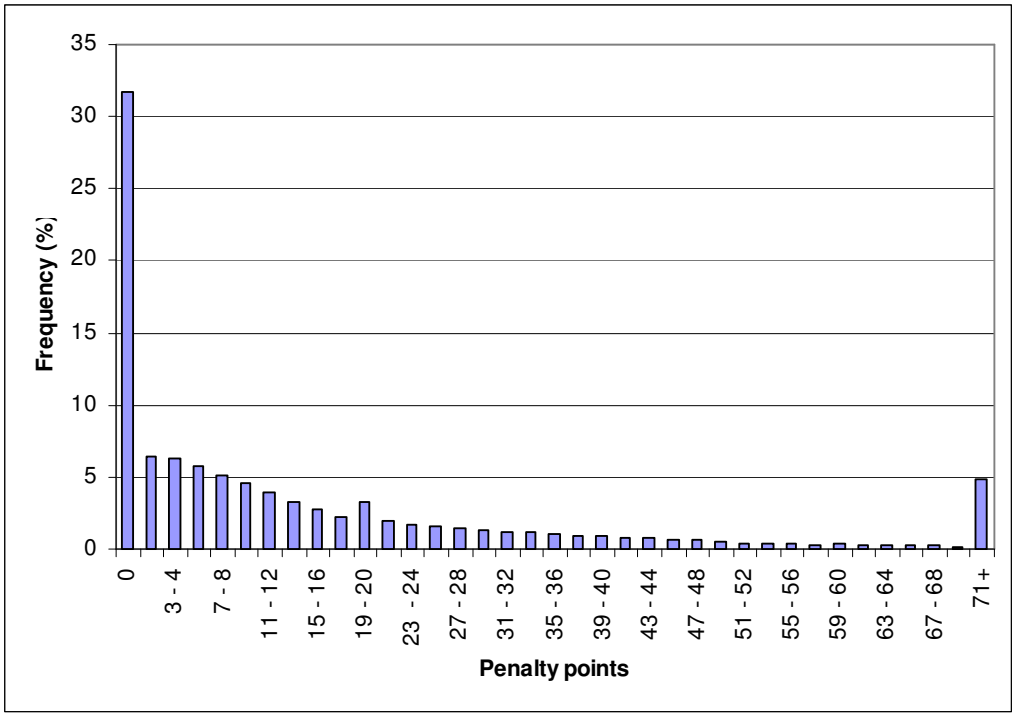


Figure 3.1c Distribution of penalty points for cross-country. Total no. records was 301318

3.4.1 Comparison of models

The computing resources required by the data augmentation method were extremely small compared to the bivariate methods. Overall fitting the data augmentation model took approximately 7 hours, compared to approximately 4 hours for each of the 66 bivariates (i.e. 264 hours in total). All runs were performed on a computer with a 2.4GHz clock speed. The bivariate analyses were each allowed up to 4GB of memory and the data augmentation method up to 8GB. The differences in the (co)variances and solutions produced by the three methods were used to compare models.

3.4.2 Phenotypic variance

As the trait was a normal score (mean 0, standard deviation 1), the phenotypic variance is expected to be less than or equal to 1, depending on class sizes. Estimates of the total phenotypic

variance for the three methods are shown in Tables 3.2 – 3.4. Estimates from all methods were close to the anticipated value of 1. Estimates from the data augmentation method had a greater range over the disciplines, compared to the bivariate methods, and tended to be higher within the more advanced grades.

3.4.3 Heritabilities

Heritability estimates for the twelve traits, as estimated by the different methods are given in Tables 3.2 – 3.4 for the individual disciplines.

Irrespective of the method of analysis, heritabilities were significantly greater than zero for all traits excluding the higher grades in cross-county. Show jumping showed the greatest heritability, followed by dressage and then cross-country.

From Method 1, the advanced grade of all disciplines had the highest heritability (Tables 3.2 – 3.4). This was particularly true for dressage (17.1%) and show jumping (22.6%). The remaining grades of dressage were very consistent (8.0 to 9.0%). For show jumping, pre-novice grade had the lowest heritability (10.6%). Note that the higher heritability estimates were associated with both higher phenotypic and higher genetic variance.

Comparing Methods 2a and 2b, differences were slight, with the greatest absolute difference in heritability estimates being 1.5%, and since the standard error of the estimate was 0.9%, this was likely to have been a chance occurrence. The standard errors were taken as averages over the bivariate analyses. Given this small difference, only Method 2a will be described more fully.

The pattern of magnitudes of heritability across disciplines and grades was similar to Method 1, with, show jumping having the highest heritability (8.9 to 16.2%), followed by dressage (7.2 to

9.0%) and then cross-country (0.3 to 1.4%) and estimates of heritability tending to be greatest in advanced grade.

3.4.4 Repeatabilities

Repeatability estimates, which are the proportion of phenotypic variance explained by the horse genetics and the horse's permanent environment combined, represent the upper limit to the heritability. These are also given in Tables 3.2 – 3.4. Methods 2a and b were in good agreement. Estimates from Method 2a were highest for dressage, 24.1 to 29.8%, followed by show jumping 15.8 to 21.0%, with cross-country the lowest again (8.5 to 9.8%).

Whilst the pattern was similar for Method 1 there were differences: for dressage at advanced grade, the estimate was lower compared to the bivariates, for show jumping advanced and intermediate grades, the estimates were higher, and for all grades of cross-country, estimates were slightly higher. However, there is no evidence to suggest that these differences are statistically significant.

3.4.5 Horse's permanent environmental variance

The proportion of phenotypic variance explained by the permanent environment of the horse is given in Tables 3.2 – 3.4. The permanent environmental variance is the variance due to environmental effects that have consistently influenced the horse's performance, such as long-term training, stabling, any early life influences and also in this analysis maternal effects which cannot be identified as dams were unknown. Estimates were again very similar between Methods 2a/b. For Method 2a the permanent environment accounted for 16.9 to 20.9 % of phenotypic variance in dressage compared to 4.9 to 7.5 % in show jumping and 6.6 to 8.7 % in cross-country. The large estimate for the influence of permanent environment on dressage

performance compared to show jumping explains why the repeatability is estimated to be greater for dressage yet its heritability appeared smaller.

For Method 1, estimates were consistent with Methods 2a/2b, with the exception of advanced grades in dressage and show jumping, which were noticeably smaller. For dressage, estimates ranged from 8.8 to 18.9 % and for show jumping from 16.9 to 24.6 %.

3.4.6 Rider variances, as a proportion of total variance

The estimates of variance due to the rider were very similar when estimated by Methods 2a and 2b. The maximum absolute difference in the estimates was 2.7% of total phenotypic variance (standard error on rider variance was 0.6%). The proportion of variance due to the rider, as estimated by Method 2a was 22.3 to 29.1% for dressage, 8.4 to 11.3% for show jumping and 9.2 to 12.0% for cross-country. Estimates by Method 1 tended to be greater in magnitude than Method 2a/b estimates.

Table 3.2 Functions of variance components for dressage as estimated by the three methods

Method		DA		DI		DN		DP	
		Est.	**s.e.**	**Est.**	**s.e.**	**Est.**	**s.e.**	**Est.**	**s.e.**
1	**Total phenotypic**	*1.098*		*1.139*		*1.074*		*1.048*	
	h^2	0.171***		0.083***		0.080***		0.090***	
	Repeatability	0.259***		0.272***		0.247***		0.246***	
	Permanent environment	0.088*		0.189***		0.167***		0.156***	
	Rider	0.372***		0.289***		0.260***		0.251***	
2a	**Total phenotypic**	*0.987*	*0.022*	*1.036*	*0.012*	*0.993*	*0.008*	*1.011*	*0.007*
	h^2	0.090*	0.045	0.076***	0.017	0.072***	0.012	0.076***	0.01
	Repeatability	0.298***	0.013	0.273***	0.006	0.241***	0.005	0.246***	0.004
	Permanent environment	0.209***	0.043	0.198***	0.017	0.169***	0.011	0.170***	0.01
	Rider	0.291***	0.019	0.244***	0.010	0.223***	0.007	0.232***	0.006
2b	**Total phenotypic**	*0.969*		*1.065*		*1.015*		*1.042*	
	h^2	0.090*		0.081***		0.080***		0.087***	
	Repeatability	0.307***		0.284***		0.242***		0.242***	
	Permanent environment	0.217***		0.203***		0.163***		0.155***	
	Rider	0.275***		0.246***		0.233***		0.251***	

DA =dressage advanced, DI=dressage intermediate, DN=dressage novice, DP=dressage prenovice
* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$; for figures in italics, the significance is not determined.

Table 3.3 Functions of variance components for show jumping as estimated by the three methods

Method		SJA		SJI		SJN		SJP	
		Est.	**s.e.**	**Est.**	**s.e.**	**Est.**	**s.e.**	**Est.**	**s.e.**
1	**Total phenotypic**	*1.044*		*0.997*		*0.940*		*0.866*	
	h^2	0.226***		0.126***		0.134***		0.106***	
	Repeatability	0.246***		0.207***		0.197***		0.169***	
	Permanent environment	0.019		0.081***		0.063***		0.063***	
	Rider	0.140***		0.141***		0.126***		0.092***	
2a	**Total phenotypic**	*0.947*	*0.015*	*0.908*	*0.008*	*0.863*	*0.006*	*0.841*	*0.004*
	h^2	0.162***	0.040	0.095***	0.014	0.094***	0.011	0.089***	0.009
	Repeatability	0.211***	0.011	0.169***	0.005	0.169***	0.004	0.150***	0.004
	Permanent environment	0.049	0.036	0.074***	0.013	0.075***	0.009	0.062***	0.007
	Rider	0.113***	0.010	0.113***	0.006	0.091***	0.004	0.084***	0.003
2b	**Total phenotypic**	*0.932*		*0.937*		*0.902*		*0.855*	
	h^2	0.155***		0.090***		0.107***		0.104***	
	Repeatability	0.210***		0.183***		0.180***		0.158***	
	Permanent environment	0.054		0.093***		0.073***		0.054***	
	Rider	0.105***		0.120***		0.111***		0.088***	

SJA=show jumping advanced, SJI=show jumping intermediate, SJN=show jumping novice, SJP=show jumping prenovice
* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$; for figures in italics, the significance is not determined.

Table 3.4 Functions of variance components for cross-country as estimated by the three methods

Method		XCA		XCI		XCN		XCP	
		Est.	**s.e.**	**Est.**	**s.e.**	**Est.**	**s.e.**	**Est.**	**s.e.**
1	**Total phenotypic**	*1.062*		*1.058*		*0.999*		*0.819*	
	h^2	0.039		0.016**		0.023***		0.018***	
	Repeatability	0.112***		0.112***		0.104***		0.101***	
	Permanent environment	0.073***		0.096***		0.081***		0.083***	
	Rider	0.181***		0.161***		0.144***		0.101***	
2a	**Total phenotypic**	*0.956*	*0.015*	*0.974*	*0.008*	*0.960*	*0.006*	*0.800*	*0.004*
	h^2	0.027	0.021	0.003	0.005	0.013**	0.005	0.014***	0.004
	Repeatability	0.093***	0.008	0.091***	0.004	0.098***	0.003	0.085***	0.003
	Permanent environment	0.066**	0.022	0.087***	0.007	0.085***	0.006	0.071***	0.004
	Rider	0.120***	0.011	0.118***	0.007	0.117***	0.005	0.092***	0.003
2b	**Total phenotypic**	*0.955*		*0.982*		*0.974*		*0.803*	
	h^2	0.032		0.008		0.021***		0.018***	
	Repeatability	0.095***		0.096***		0.103***		0.086***	
	Permanent environment	0.064**		0.088***		0.082***		0.069***	
	Rider	0.117***		0.119***		0.124***		0.093***	

XCA=cross-county advanced, XCI=cross-country intermediate, XCN=cross-country novice, XCP=cross-country prenovice
* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$; for figures in italics, the significance is not determined.

3.4.7 Genetic correlations

Genetic correlations are presented in Table 3.5. Correlations within grades were all significant. For dressage, these ranged from 0.783 to 0.946, for show jumping, from 0.803 to 0.954, and for cross-country, from 0.358 to 0.698. There was a clear pattern that correlations between sequential grades were higher than those between more distant grades. Results from Methods 2a and 2b were very similar. For both methods, genetic correlations between the grades were significantly different from zero for dressage and show jumping, positive and high (for dressage 0.59 to 0.99, for show jumping 0.74 to 0.99). For cross-country, there is a need for greater caution as genetic correlations between the grades were only significantly different from zero between novice and pre-novice, with an estimate of 0.75. This provides evidence that for dressage and show jumping, and at least within lower grades of cross-country, within the discipline, the same loci (or loci in linkage disequilibrium) are predominantly responsible for performance at the different grades.

For Method 1, correlations between disciplines were largely significant. Between show jumping and dressage, significant correlations ranged from 0.011 to 0.385, but estimated correlations between cross-country and other disciplines appeared more heterogeneous in sign ranging from -0.163 to 0.093 for dressage. By contrast, from Methods 2a and 2b between discipline correlations were largely not significant, but also of varied sign and magnitude. Correlations between the lower grades of dressage and show jumping were statistically significant, positive but low, and a low significant correlation between show jumping and cross-country pre-novice and novice was found. Estimates for dressage and cross-country were indicative of there being negative correlations of moderate magnitude. For Method 2b, comparing the genetic correlation

matrix before and after bending (considering only estimates that were significantly different from zero in the positive definite matrix) the average absolute change was 0.054.

3.4.8 Horse and rider correlations

Correlations for the horse (representing the permanent environment and 3* the sire genetic component) between grades, within disciplines were all significant and generally high (Table 3.5). Rider correlations are in Table 3.6.

From Model 1, correlations between disciplines were largely significant, although low/ moderate. Correlations between show jumping and cross-country were slightly higher than between other disciplines. Rider correlations were all significant. Within discipline correlations were very high. Between discipline correlations were moderate/ high, although slightly lower between dressage and cross-country compared to the other disciplines.

By Method 2b, horse correlations within disciplines were 0.64 or more for dressage, but showed a greater range of 0.36 – 0.96 for show jumping and 0.10 – 0.80 for cross-country. The correlation structures were clearly banded, with high or moderate correlations observed between adjacent grades, lower correlations between grades once removed and lowest between Pre-Novice and Advanced. Significant correlations between disciplines were positive but low. In general, rider correlations, both within and between disciplines, were significant. Within discipline correlations were high, and between discipline correlations generally moderate.

Table 3.5a. Horse permanent environment (below diagonal) & genetic (above diagonal) correlation matrices for the twelve traits, as estimated by methods 1 & 2a

		DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
1	**DA**		0.937*	0.846*	0.783*	0.001	0.197*	0.303*	0.258*	-0.017*	-0.071*	0.093*	-0.052*
	DI	0.901*		0.921*	0.869*	0.011*	0.259*	0.333*	0.313*	-0.069*	-0.075*	-0.016*	-0.057*
	DN	0.730*	0.920*		0.946*	0.085*	0.312*	0.356*	0.378*	0.019*	-0.058*	-0.073*	0.021*
	DP	0.609*	0.782*	0.900*		0.084*	0.297*	0.325*	0.385*	-0.061*	-0.098*	-0.163*	0.061*
	SJA	0.395*	0.384*	0.351*	0.354*		0.880*	0.843*	0.803*	0.082*	0.319*	0.292*	0.210*
	SJI	0.178*	0.179*	0.128*	0.106*	0.914*		0.954*	0.924*	-0.057*	0.288*	0.280*	0.249*
	SJN	0.060*	0.096*	0.131*	0.094*	0.757*	0.922*		0.945*	0.014*	0.313*	0.405*	0.263*
	SJP	0.281*	0.213*	0.199*	0.231*	0.527*	0.685*	0.833*		0.028*	0.312*	0.342*	0.378*
	XCA	0.279*	0.243*	0.174*	0.189*	0.571*	0.316*	0.316*	0.147*		0.603*	0.422*	0.358*
	XCI	0.191*	0.299*	0.234*	0.138*	0.435*	0.380*	0.380*	0.159*	0.858*		0.698*	0.492*
	XCN	-0.027*	0.157*	0.219*	0.145*	0.231*	0.447*	0.447*	0.305*	0.457*	0.773*		0.550*
	XCP	0.241*	0.186*	0.216*	0.287*	0.172*	0.259*	0.259*	0.466*	0.254*	0.305*	0.623*	
2	**DA**		0.937*	0.846*	0.783*	0.001	0.197*	0.303*	0.258*	-0.017*	-0.071*	0.093*	-0.052*
	DI	0.901*		0.921*	0.869*	0.011*	0.259*	0.333*	0.313*	-0.069*	-0.075*	-0.016*	-0.057*
	DN	0.730*	0.920*		0.946*	0.085*	0.312*	0.356*	0.378*	0.019*	-0.058*	-0.073*	0.021*
	DP	0.609*	0.782*	0.900*		0.084*	0.297*	0.325*	0.385*	-0.061*	-0.098*	-0.163*	0.061*
	SJA	0.202*	0.093*	0.028	0.040		0.902*	0.895*	0.740*	-0.358	*-0.562*	-0.002	0.190
	SJI	0.063	0.105*	-0.009	-0.012	0.963*		0.985*	0.948*	0.025	*0.000*	0.154	0.225
	SJN	-0.041	-0.004	0.113*	0.050*	0.706*	0.912*		0.951*	-0.068	0.071	0.273*	0.142
	SJP	-0.060	-0.060	0.068*	0.139*	0.362*	0.645*	0.842*		-0.317	-0.229	0.208	0.260*
	XCA	0.184*	0.172*	-0.027	0.013	0.407*	0.235*	-0.009	0.044		*0.989*	*0.897*	0.561
	XCI	0.029	0.170*	0.023	-0.049	0.324*	0.319*	0.116*	0.021	0.800*		0.665	0.046
	XCN	-0.078	0.032	0.136*	0.052*	0.148*	0.051	0.339*	0.118*	0.264*	0.691*		0.753*
	XCP	0.098	0.017	0.074*	0.161*	0.054	-0.031	0.049	0.273*	0.099	0.283*	0.493*	

Correlations other than those in italics were significantly greater than zero ($P < 0.05$); figures in italics were estimated at the boundary and no standard errors were available.

Table 3.5b. Horse permanent environment (below diagonal) & genetic (above diagonal) correlation matrices for the twelve traits, as estimated by method 2b

2b	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
DA		0.914*	0.788*	0.677*	-0.439	-0.308	-0.135	-0.074	*-0.078*	*-0.217*	0.163	-0.152
DI	*0.950*		0.899*	0.896*	-0.268	-0.064	0.085	0.186	-0.366	-0.220	-0.020	-0.146
DN	0.745*	0.876*		0.948*	-0.188	0.071	0.117	0.228*	-0.296	-0.174	-0.230	-0.098
DP	0.642*	0.717*	0.911*		-0.072	0.169	0.224*	0.347*	-0.504	-0.152	-0.298	-0.065
SJA	0.170*	0.109*	0.013	0.038		0.832*	0.826*	0.756*	-0.163	*-0.234*	0.002	0.157
SJI	0.078	0.072*	0.011	-0.016	0.924*		0.961*	0.955*	-0.073	*-0.039*	0.095	0.174
SJN	-0.044	0.002	0.100*	0.058*	0.671*	0.853*		0.982*	-0.095	-0.082	0.248	0.152
SJP	-0.062	-0.049	0.071*	0.139*	0.351*	0.597*	0.825*		-0.171	-0.122	0.172	0.172
XCA	0.178*	0.170*	-0.031	0.017	0.368*	0.252*	-0.009	0.025		*0.445*	*0.704*	0.451
XCI	0.038	0.145*	0.030	-0.051	0.315*	0.272*	0.134*	0.019	0.798*		0.354	0.188
XCN	-0.077	0.032	0.133*	0.052*	0.119	0.089*	0.300*	0.130*	0.257*	0.672*		0.378*
XCP	0.086	0.020	0.067*	0.167*	0.041	-0.031	0.062*	0.262*	0.104	0.266*	0.486*	

* = Correlation significantly greater than zero ($P < 0.05$); figures in italics were estimated at the boundary and no standard errors were available.

Table 3.6a. Rider correlation matrices for the twelve traits, as estimated by methods 1 &2a

		DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
1	**DA**	0.973	0.939	0.896	0.587	0.635	0.641	0.623	0.504	0.467	0.478	0.490
	DI		0.975	0.943	0.612	0.689	0.709	0.697	0.485	0.518	0.541	0.581
	DN			0.982	0.635	0.717	0.735	0.723	0.453	0.492	0.522	0.579
	DP				0.615	0.717	0.735	0.730	0.403	0.489	0.528	0.610
	SJA					0.912	0.892	0.860	0.669	0.615	0.635	0.614
	SJI						0.980	0.963	0.574	0.698	0.692	0.697
	SJN							0.983	0.524	0.662	0.706	0.706
	SJP								0.457	0.655	0.707	0.738
	XCA									0.802	0.706	0.588
	XCI										0.938	0.841
	XCN											0.931
2a	**DA**	*0.968*	0.906	0.807	0.792	0.740	0.606	0.539	0.610	0.355	0.352	0.324
	DI		0.974	0.915	0.763	0.758	0.665	0.645	0.543	0.387	0.397	0.415
	DN			0.972	0.653	0.733	0.601	0.611	0.574	0.463	0.441	0.458
	DP				0.513	0.644	0.534	0.549	0.438	0.426	0.404	0.475
	SJA					0.968	0.893	0.790	0.566	0.477	0.468	0.378
	SJI						0.994	0.906	0.614	0.571	0.527	0.429
	SJN							0.905	0.620	0.544	0.568	0.444
	SJP								0.619	0.611	0.548	0.609
	XCA									0.854	0.765	0.627
	XCI										0.919	0.683
	XCN											0.783

Correlations other than those in italics were significantly greater than zero ($P < 0.05$); figures in italics were estimated at the boundary and no standard errors were available.

Table 3.6b. Rider correlation matrices for the twelve traits, as estimated method 2b

2b	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN
DA	*0.972*	0.934	0.837	0.632	0.642	0.595	0.520	0.487	0.297	0.315	0.292
DI		0.987	0.938	0.624	0.673	0.628	0.597	0.474	0.352	0.365	0.390
DN			0.975	0.556	0.636	0.596	0.585	0.516	0.426	0.423	0.436
DP				0.470	0.579	0.539	0.572	0.430	0.414	0.411	0.474
SJA					0.974	0.951	0.768	0.445	0.377	0.404	0.301
SJI						0.986	0.859	0.528	0.481	0.472	0.379
SJN							0.911	0.610	0.571	0.564	0.475
SJP								0.607	0.579	0.560	0.627
XCA									0.891	0.833	0.616
XCI										0.964	0.689
XCN											0.799

Correlations other than those in italics were significantly greater than zero ($P < 0.05$); figures in italics were estimated at the boundary and no standard errors were available.

3.4.9 Comparison of correlation estimates between Method 2a and Method 1

Figure 3.2 shows a more detailed comparison of methods restricted to estimates that were within discipline and statistically different from 0 (based on standard errors). The genetic correlations, once away from 1, showed more scatter, probably due to their greater sampling error. The horse correlations, comprising both the genetic and permanent environment, were broadly in agreement between the methods, with no evidence of consistent bias. For the rider correlations, again there was no evidence of a consistent bias between methods.

Between discipline correlations were different (not shown). There was a consistent trend for Method 1 to estimate genetic correlations or horse correlations that were larger than for Method 2a (with the comparison restricted to those that were judged to be statistically significant). The same was true for rider correlations, with the exception of correlations including some of the more advanced grades, where estimates by Method 2a were often higher than Method 1.

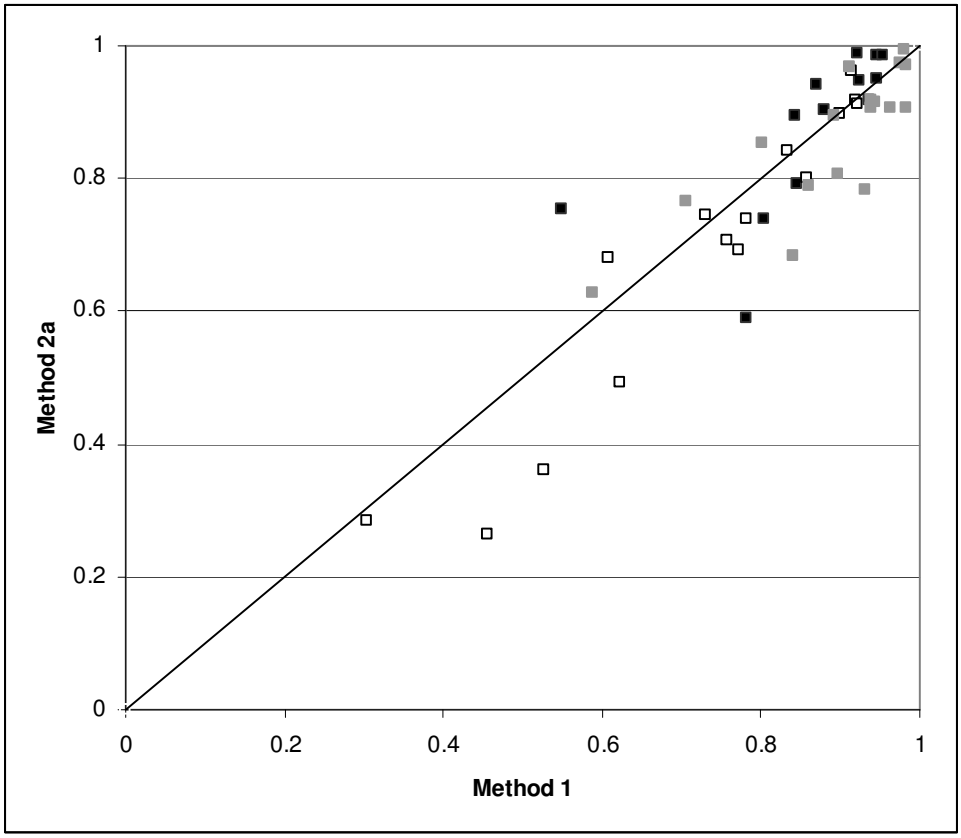


Figure 3.2. Estimates of correlations within grades for the genetic (black), horse (white) and rider (grey) effects plotted for method 1 versus 2a
Only correlations where standard errors indicated that values were significantly different from zero are included

3.4.10 Fixed effects of gender and age

For the effects of gender, for all models, there were generally significant gender effects for dressage, but not for the other disciplines. For dressage, stallions and geldings performed significantly better than mares in most grades. For show jumping, estimates were in general not significant, but stallions performed better than mares and mares performed better than geldings. For cross-country, by Method 1, geldings performed better than mares, and mares better than stallions. For the Methods 2a & b, there were no discernable differences for cross-country.

For all models and all traits there was a significant quadratic association between performance and age, indicating a curvi-linear increase in performance with age. Figures 3.3 – 3.5 illustrate the change in performance with age, using example results from the bivariate analyses. As might be predicted, performance in the more advanced grades peaked at a later age. For all disciplines, similar patterns were observed, although between disciplines peak performance between grades varied and was least pronounced in show jumping. .

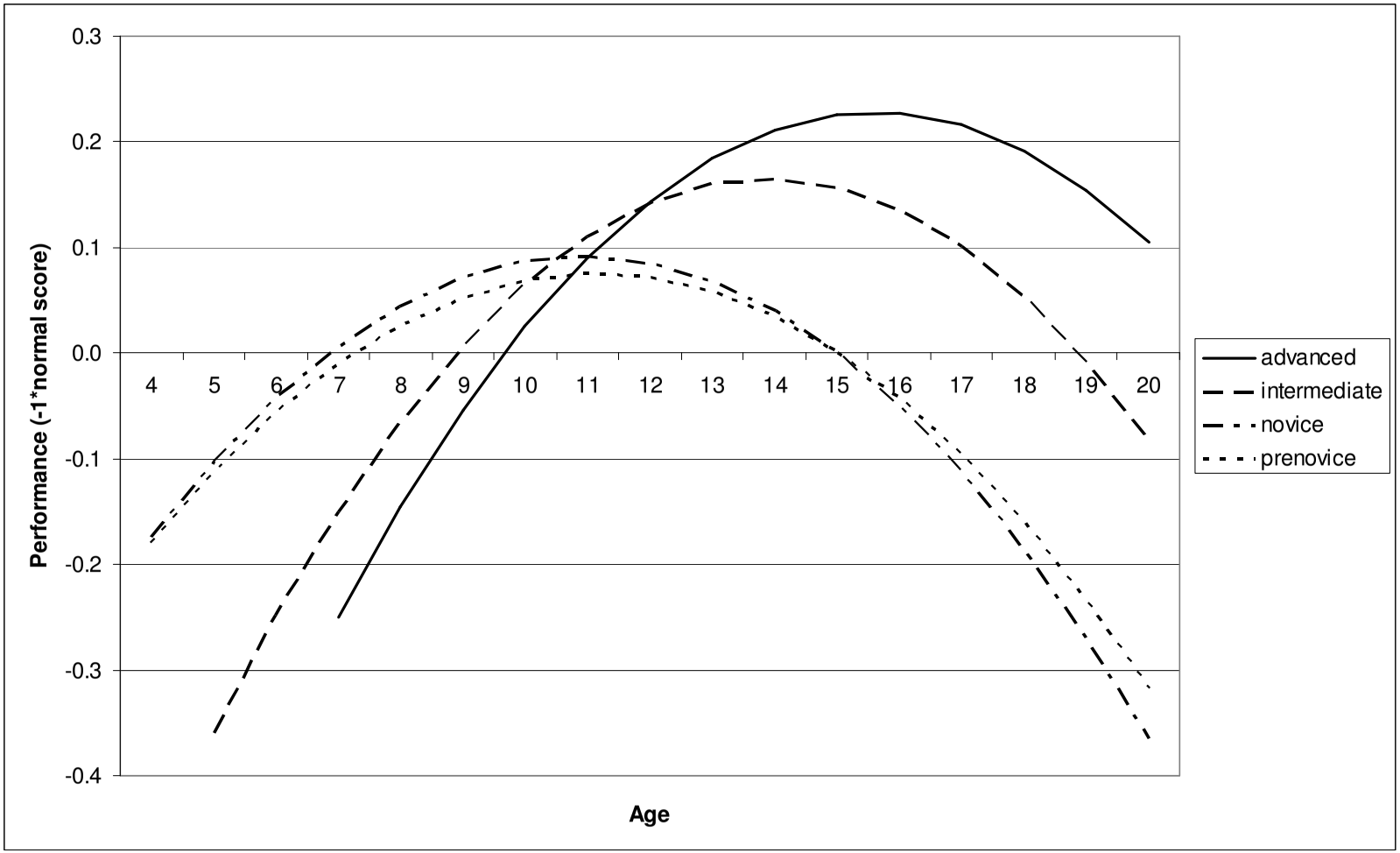


Figure 3.3 Performance with age for each grade of dressage. Horses >= 20 years were grouped.

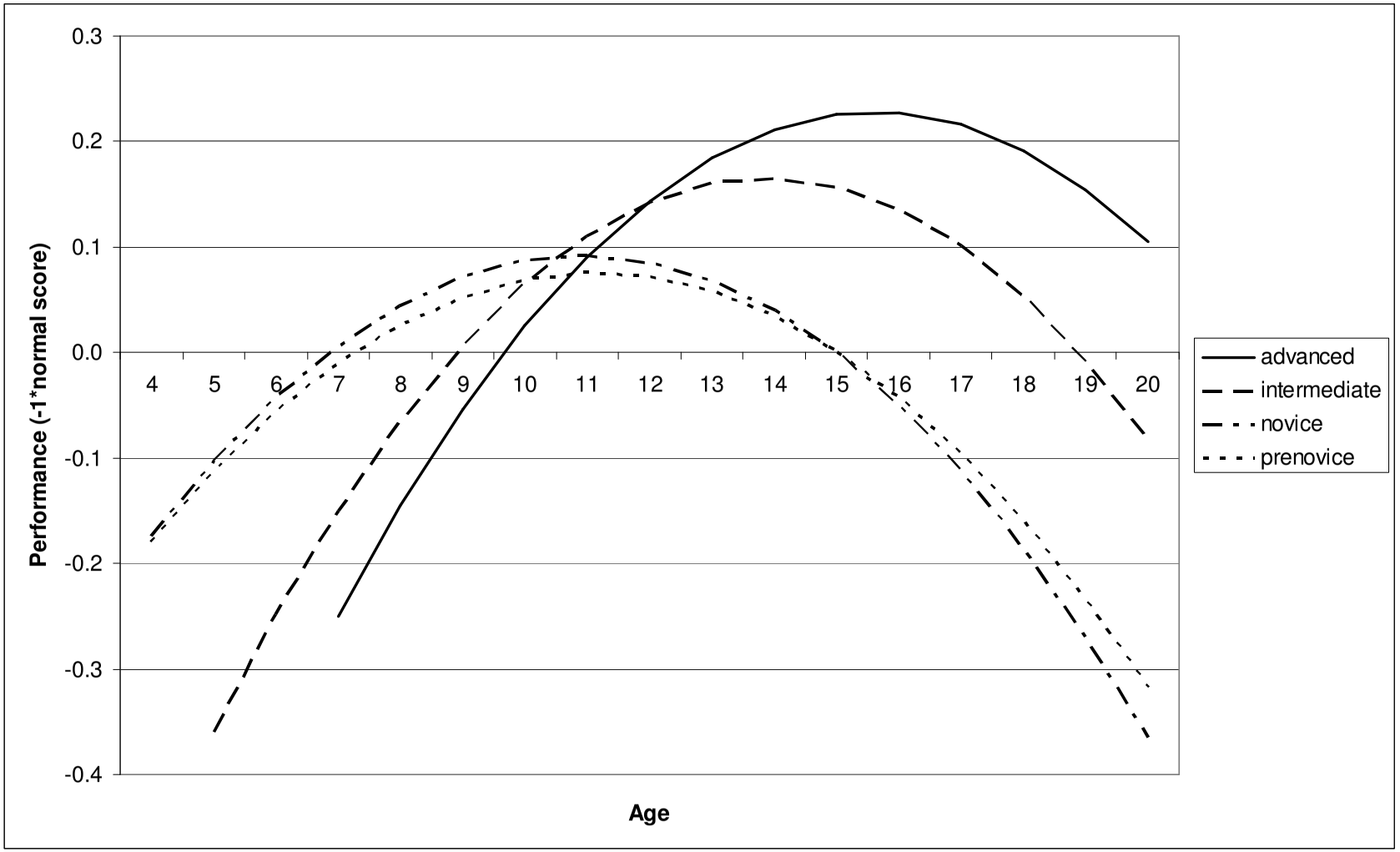


Figure 3.4 Performance with age for each grade of show jumping. Horses >= 20 years were grouped.

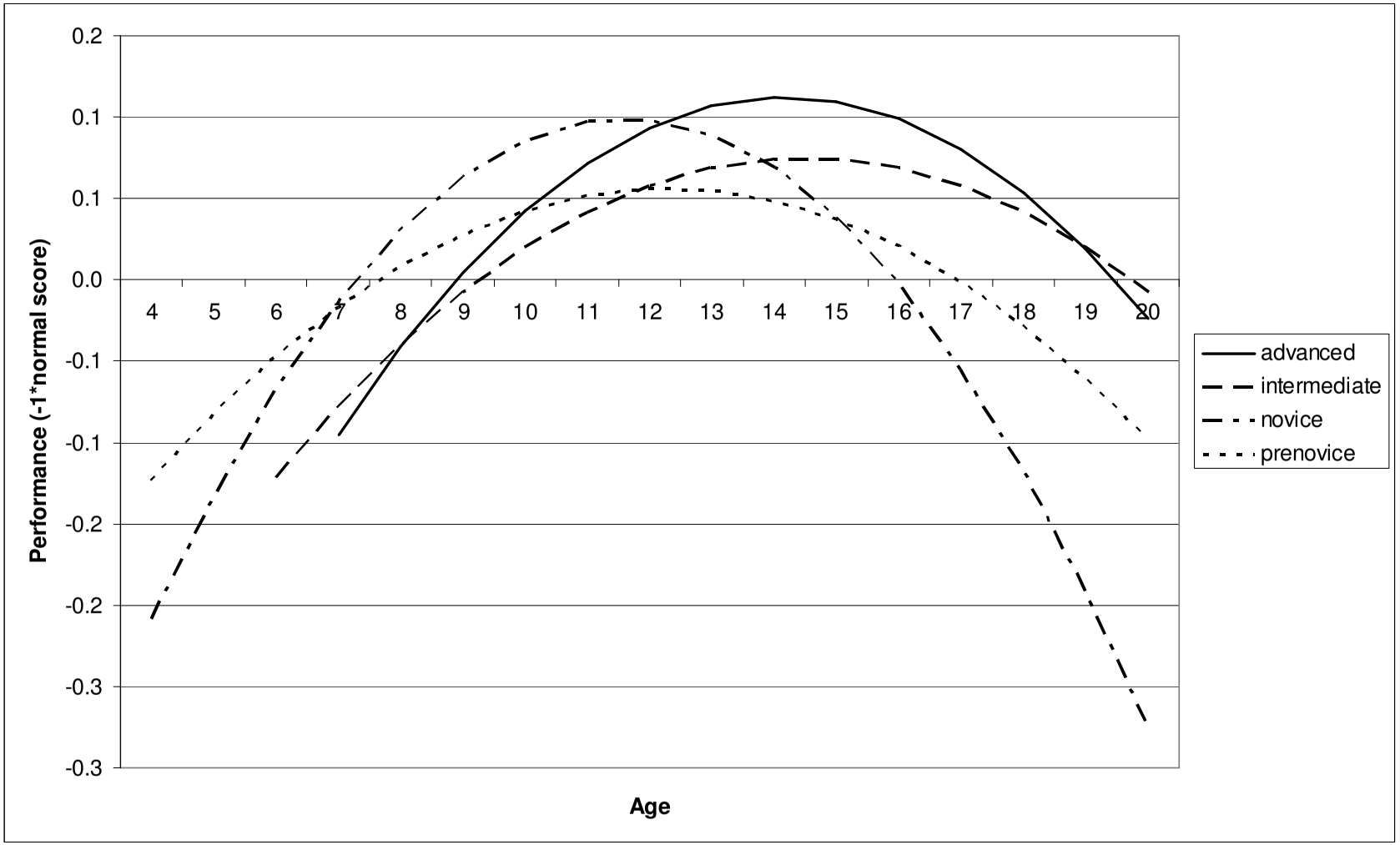


Figure 3.5 Performance with age for each grade of cross-country. Horses >= 20 years were grouped.

3.5 Discussion

There is a demand for tools to estimate (co)variance matrices for a large number of traits in an acceptable time frame. The current test day model used by the dairy cattle industry in the UK requires a 27*27 (co)variance matrix, and other industries may desire even larger matrices. For example a BLUP with 100 traits would be ideal for the silvicultural objectives of Skogforsk (the Forestry Research Institute of Sweden) (B. Andersson personal communication December 2010). A rapid analysis would enable easy re-estimation of (co)variance matrices, rather than the current situation where in many genetic evaluation systems, due to computing overheads, the (co)variance components are not re-estimated regularly and BLUP EBV are being produced based on variance components compiled using *ad hoc* methods from old data. Therefore any method offering an advance in such methodology would have important applications.

We compared three methods of producing large (co)variance matrices for genetic evaluations, using a dataset of results from eventing competitions in the UK. Two methods that combined smaller sub-matrices from bivariate analyses were investigated: i) Method 2a, a novel method based on Wall *et al.* (2005) that generates samples of each bivariate distribution to perform a multivariate analysis and obtain a positive definite 12*12 (co)variance matrix and ii) Method 2b, spectral decomposition of the matrix of mean results replacing negative eigenvalues with positive values to obtain a positive definite (co)variance matrix. The latter is likely to be relatively well applied in practice. The two 'bivariate' methods, differing only in the way that the bivariate matrices were combined, gave, in this case, very similar results for functions of variance components. Method 1 was the novel data augmentation function in ASReml. The multivariate nature of Method 1 will result in greater accuracy (with the increase in accuracy

determined by the absolute differences between the residual and genetic correlations between traits).

(Co)variance estimates and functions of variance components produced by the three methods were in general similar. Correlation estimates were more often significant for Method 1, which may be partially due to the fact that standard errors for these values were approximate values.

As the trait was a normal score (mean 0, standard deviation 1), the variance of each trait would be expected to be less than or equal to 1, depending on class sizes. In the data augmentation analysis, the increase in phenotypic variance with grade within a discipline could be explained by the fact that Method 1 was a 12-trait analysis, with information on performance at all grades included and thus the analysis at least partly accounts for selection between grades, as well as producing more accurate variance component estimates. It would also be feasible to obtain phenotypic variances greater than 1 when analysing heritable traits with substantial inbreeding with a pedigree of many generations in depth. However, in this analysis it was only possible to use a sire model which makes no allowances for inbreeding.

No formal comparison of the fits of the models could be performed. The three methods produced largely consistent results. Methods 2a and 2b are suitable for implementation and will produce large, positive definite (co)variance matrices, however require large computing resources. Method 1 produced results which were generally consistent with Methods 2a & b. Method 1 enables the production of large (co)variance matrices in a computationally efficient way, and harvests the full benefits of a multivariate analysis. Differences in variance component estimates between Methods 1 and 2a/b may be due to the increased accuracy of Method 1, a result of the multivariate nature of the analysis.

The heritability estimates for the dressage phase of the competition are comparable with estimates for the individual competition discipline in GB, which were 0.15 (s.e. 0.018) when breed is not accounted for (Stewart *et al.,* 2010), and also with international estimates. For example, heritabilities ranging from 0.10 to 0.20 have been estimated in various breeds (see Huizinga and Van Der Meij*,* 1989*;* Ricard *et al.,* 2000; Ducro *et al.,* 2007; Janssens, 2008; Olsson *et al.,* 2008). Similarly, heritability estimates for show jumping are comparable to international estimates for competition data, ranging from 0.10 (Hanoverian, Trakehner, Oldenburg) through 0.20 in Dutch Warmbloods (Huizinga and Van Der Meij*,* 1989) to 0.28 in the Swedish Warmblood (Olsson, 2006; Janssens, 2008).

The genetic variation detected in the analysis for cross-country was low and to some extent it may be masked by greater environmental variance. Improved pedigree data may help to elucidate the genetic variation further. However, our results indicate that in a selection program, the greatest genetic gain can be achieved in show jumping, followed by dressage, and that little gain will be made in cross-country. Therefore, a breeding program designed to select for eventing competition performance in GB will be best suited to select for the show jumping or dressage phases of the competition. Even if this were considered as a phenotypic selection rather than a selection for breeding, performance at lower grades was a poorer predictor of performance at high grade for cross-country than for show jumping or dressage. At present in the UK, the situation in practice is directly contrary to this, as breeders of eventing horses tend to consider cross-country performance as the primary selection aim (J. Rogers personal communication May 2010). This traditional practice is derived from the origins of the discipline which originated in the Military and was dominated by cross-country. Greater weighting was placed on cross-country, with requirements for dressage and show jumping lower. The dressage phase was aimed at increasing the control the rider has over the horse, and the show jumping

phase was used to test the recovery of the horse after a cross-country competition. More recently, due to international influences, the regulation of the discipline has changed. A lower weight has been given to cross-country which is judged to be too dangerous and less and less conforming to animal welfare regulations. This has changed the type of horses needed to compete in this discipline.

Within the disciplines of dressage and show jumping, correlations between the grades were high, indicating that the same loci (or loci in linkage disequilibrium) are responsible for performance at the different grades, and therefore selection for performance in one grade also selects for performance in another grade. Analogous correlations between the grades for cross-country were estimated by Method 1 as significantly different from zero but more moderate; results by Methods 2a/b were harder to estimate due in part to a lack of genetic variance detected, particularly in the higher grades.

There was evidence for slight confounding between grade and age, with horses starting to compete in more advanced grades at older ages, and old horses tending to compete in the advanced grades. However, this will not have influenced the variance component estimates, or bias the estimates of fixed effects (although the latter are estimated with less accuracy).

Overall, the results from the data augmentation method indicate that selection for performance in dressage will improve performance in show jumping and that performance in show jumping will improve performance in cross-country. There was an indication that selection for dressage merit would reduce the merit for cross-country. Higher correlations between the lower grades may be due to general traits that jointly affect performance in both, i.e. an amateur horse that is athletic and with good temperament should have good ability, at lower levels of competition, in all

disciplines. However at the highest grade of competition these effects are less apparent, where more specific traits, such as the gaits and jumping have more influence on performance.

Variance components due to the horse and rider were considered as separate random effects. This was possible due to the fact that one rider often rode multiple horses, and that many horses were ridden by multiple riders. If the latter were not true, different treatments of the rider effect within the model could be considered. For instance, a hierarchical model (of horse within rider) could be used. Alternatively rider could have been included as a fixed effect.

Gender effects for show jumping and particularly dressage found indication of an advantage for stallions. Given that stallions are a selection of the male population retained for breeding, this is understandable. However, gender effects were often not significant. It was surprising that no gender effect was detected for cross-country, as there is clear evidence of an advantage for males as evidenced for racing (Entin 2008).

The analysis was restricted to a sire model, so genetic connections were limited (restricted to within half-sib groups). Although this was sufficient to proceed with the analysis, the precision of the estimates of variance components and EBVs will have been reduced, with this reduction reflected in the standard errors. Assortative mating was not accounted for, possibly inflating the sire variance component, and overestimating EBVs for superior sires (and underestimating for inferior sires). Selection and inbreeding were also not accounted for. In horse populations, the assumption of random mating rarely holds. Selection has been practiced over time and non-random mating occurs, in that superior horses tend to be mated as do inferior horses, and there may be specialised breeding for the different disciplines. Another potential problem is the

selection of horses present in the data. The consequence of this will be to reduce heritability estimates.

The breeding objective for the British sport horse is to cater for both professional elite levels of competition and riders as well as amateur riders. Traits such as athleticism, soundness, rideability and conformation suitable for competing in a range of disciplines, including dressage, show jumping and eventing are required. The data augmentation 12-trait analysis enables all traits to be considered in the analysis, and thus the accuracy of estimated breeding values is maximised. The results for the analysis suggest that there are positive genetic correlations between dressage and show jumping and show jumping and eventing, indicating that selection for one discipline will also enhance performance in another. However, correlations between eventing and dressage were sometimes negative. The British sport horse has made its name in eventing competition, and thus selection for this discipline may reduce performance in dressage.

There is little literature assessing genetic correlations between adult competition disciplines; instead data tends to be sourced from young horse tests. Internationally, estimates for genetic correlations between dressage and show jumping related traits at young horse tests are generally low and mixed (either positive or negative), although correlations between canter and jumping tend to be higher than correlations between the other gaits and jumping (Thorén Hellsten *et al.,* 2006). This study adds some evidence to this area, indicating that simultaneous selection for both dressage and show jumping may hinder genetic progress for the disciplines individually. Genetic correlations between levels of competition have been estimated however, for example Lüehrs-Behnke *et al*., (2006c) estimated correlations between grades for dressage and show jumping individually. Correlations between grades within disciplines were high, or moderate to

high for show jumping and dressage respectively, although heritability estimates for show jumping were substantially lower than in this study.

Twelve estimated breeding values for each of the discipline-grade traits are now available. These can be combined into an index for overall competition performance, with weightings depending on the breeding purpose. In breeding for elite levels of competition, performance at the more advanced grades is important. However, a main aim of sport horse breeding in the UK is to produce horses for amateur and young riders, which constitute the vast majority of the riding and competing population, and for which traits such as temperament and general ability have most influence. The accuracy of the estimated breeding values at lower and intermediate levels of competition is likely to be higher, as the accuracy of estimated breeding values is dependent on the number of records. At advanced grade a horse may compete far less than at lower grades (possibly only four times per year), resulting in fewer records and hence lower accuracies for advanced estimated breeding values.

The data augmentation function in ASReml presents a valuable tool for multivariate mixed effects models. However, care must be taken in application. As large numbers of traits are rapidly and more easily evaluated, it becomes possible for analyses to be conducted as a "fishing expedition", where hypotheses are tested in a *post hoc* manner, and where overall conclusions maybe drawn on the strongest levels of significance observed for individual traits. As larger numbers of traits are tested, the probability of one appearing significant purely by chance (type 1 error) increases. For variance component estimates, the likelihood that any one estimate is not within confidence intervals would increase. In these situations, correcting for multiple testing should be considered, so that significance thresholds are set more stringently.

3.6 Conclusion

In conclusion, this work investigated three methods of estimating large (co)variance matrices. Two of these methods, based upon bivariate analysis, are computing intensive but reliable and appropriate for implementation at present. The third method (using data augmentation) requires far fewer computing resources, is much quicker to run, and will be valuable in the future for the production of large (co)variance matrices.

4. Genetic Evaluations of Traits Recorded in British Young Horse Tests

4.1 Abstract

Tests for young sport horses were recently introduced in GB. This study characterises the young horse data, examines their suitability for genetic analysis and estimates the genetic parameters needed for breeding value prediction.

Evaluation data from 2006 – 2009 were used. This included 1887 evaluations of 1323 horses, which were evaluated for competing in dressage, show jumping, eventing, endurance or as sport ponies. Traits assessed were conformation, correctness of paces, type and temperament, athleticism and veterinary. The distributions of traits were examined and correlations of traits between disciplines, for the effect of the horse, were estimated. These indicated that traits could be assumed to be genetically identical across disciplines. Variance components were estimated for each of the 5 traits, using an animal model, where random effects were the genetic effect of the horse and the permanent environment of the horse. Bivariate analyses were performed between pairs of traits.

Mean scores for each trait in each discipline were between 8.02 and 8.24, and standard deviations were between 0.54 and 0.83. Heritabilities ranged between 20.3% for athleticism and 42.2% for type and temperament. The variance due to the horse's permanent environment ranged from approximately 25% for correctness of paces and athleticism to 51.6 % for veterinary. The genetic correlations between traits were generally high.

The young horse tests ("Futurity") recently introduced in the UK are a valuable data source for genetic evaluations. The most appropriate measure will be to combine young horse data with adult competition data to routinely estimate breeding values.

4.2 Introduction

Data from young horse tests are frequently used in genetic evaluations of sport horses internationally, either alone, or in combination with adult competition data. For instance, young horse data are used in the genetic evaluations in Sweden, The Netherlands, Belgium, Germany and France, amongst others, where 'young' may mean ages between 3 and 7 years of age (Thorén Hellsten *et al.,* 2006). The genetic evaluations are made possible by the routine conduct of trials organised by the national breeding organisations, over many years and resulting in large datasets. Data are generally collected from field tests, station tests or young horse competitions, which may last between 1 and 100 days, and can involve up to 45% of eligible horses (Thorén Hellsten *et al*., 2006). The primary aims are both genetic, to select stallions and mares for breeding, and phenotypic, to identify horses with a high potential to succeed in competition. Commonly, breeding objectives focus only on show jumping and/ or dressage, and tests examine the conformation, gaits, performance, behaviour and health of the horse.

Analogous young horse tests (Young Horse Evaluations) were introduced in the UK relatively recently, in 2002, and were developed with particular reference to the form and objectives of two Swedish tests – Young Horse Tests (YHT) and Riding Horse Quality Tests (RHQT), aimed at 3 and 4 years and 4 years respectively. The UK tests were initially aimed at 4 to 6 year olds, but recruitment proved difficult due to the introduction of young horse classes for 4 year olds by competition bodies (J. Rogers personal communication), and the accumulated data proved too small for the estimation of genetic parameters (Kearsley, 2008). Therefore a set of new young

horse tests ("Futurity") was developed, targeted primarily at 3 year old horses, but including foals to 3 year olds (4 years olds in 2007). These evolved from the Young Horse Evaluations and were established by a national breeding initiative (British Equestrian Federation's British Breeding) in 2005. There are dual aims of the Futurity tests: to identify elite horses for competition in the different disciplines of dressage, eventing and show jumping, as well as to inform breeding in the UK by providing data for genetic evaluations. Participation has grown since this scheme was introduced and now attracts approximately 900 horses per year with owners motivated by assessing their horse's future potential in competition, and obtaining gradings for marketing purposes.

Given that young horse data have been collected in the UK since 2005, a dataset of adequate size has only recently become available for use in genetic evaluations. The aim of this work therefore was to 1) characterise the young horse data, 2) investigate its suitability for genetic analysis, 3) estimate the genetic parameters needed for breeding value prediction and 4) predict breeding values for all horses in the pedigree.

4.3 Material and Methods

4.3.1 Data

Test data for 1323 horses of all sexes were obtained from the Futurity database for the years 2005 – 2009. Data from 2005 were excluded, as this was a pilot year, with only 72 records. Horses eligible for the Futurity tests must be British bred (foaled in the UK, or to a mare that is usually resident in UK) and can be aged from a few weeks to 3 years old (4 years in 2007). Horses are entered by owners/ breeders, and it is acceptable for them to be evaluated in multiple years. Evaluations are conducted in a single day at events held over the summer months at

various locations across the UK. The data collected represent a total of 35 evaluation days, across 15 locations, with a range of 7 to 79 horses attending per day.

At an evaluation, horses are judged by a team of up to four judges and a veterinary practitioner. Horses are judged within an arena, without rider, where they are led through various tests. They are examined for their potential future performance in one of the specific disciplines - show jumping, dressage, eventing, endurance or sports ponies (the latter defined as mature height less than 148cm and at least one parent under 148cm). The same traits are examined for each discipline: conformation, correctness of paces, type and temperament, athleticism, and veterinary scores. The precise definition of traits differs between disciplines, for example, if the same horse is evaluated for both show jumping and dressage, it might well receive different grades for paces, as the required paces will differ between the disciplines. Subjective scores from 1 (very poor) to 10 (excellent) are awarded for each trait, in increments of 0.25, and an overall grade summarising all the traits is also awarded. Horses are evaluated for the one discipline for which they are entered by the owner (i.e. they are not, as standard, examined for both dressage and show jumping, as is performed in most European evaluations). In the first years of the futurity (2006 – 2009) horses could be entered for more than one discipline; as of 2010 the rules changed so that horses may only be entered for one discipline. Judges may advise that a horse is evaluated for an alternative discipline if thought appropriate, but whilst the occurrence of this was frequent in the early years, the frequency has decreased over time.

The traits have varied over the years. In 2006, no veterinary trait was recorded and in 2008 ‘type’ became the more general ‘type and temperament’. Both conformation and veterinary were split into two better defined traits in 2008, as was correctness of paces in 2009; in the analysis the two recordings were averaged to produce a single score. The trait definitions have become

more precise over time. For instance, the 2009 definition for type and temperament in eventing was “Suitability of type and temperament to include attention, confidence, expression, efficiency of movement and harmony specifically in relation to intended discipline and relative to age. Is intelligent, willing and honest in his/her attitude to jumping, evaluates the jump and if he/she makes a mistake – seeks to tackle it differently next time. Should give the impression of being trainable”. The veterinary mark (for the same year and discipline) was awarded for the conformation of limbs, hooves, musculo-skeletal and other biomechanical factors which could influence future performance, and assessed the horse’s potential to stay sound and free from injury. Good conformation required a rectangular proportional build with horizontal back and proportional legs, a supple poll and head/neck connection with clean throat latch, a long arched neck with muscling to top line, and a strongly built and muscled back and loin. Correctness of paces was judged individually for trot and walk. Walk was an active gait with impulsion and was supple, free in the shoulder and elbow and showed a noticeable over track. Trot was an active gait with impulsion and was supple with balance and self carriage, with a “rounder” action more than a “toe flick” and was rhythmical. Athleticism assessed the canter and gallop, and for three-year olds the jump was assessed.

Also included in the extracted data were the date and location of evaluation, date of birth, age, gender, colour of horse, owner, breeder and handler. Pedigree data were available with the sire, dam and dam’s sire identified. Four year old horses were evaluated in 2007 only, and accounted for only 6 records.

4.3.2 Statistical Methods

4.3.2.1 Phenotypic analysis

The distributions of scores from the 5 traits: 1) athleticism 2) conformation 3) correctness of paces 4) type and temperament and 5) veterinary were examined. All scores were greater than or equal to 4, with the exception of one value of 2.5. All other scores for the individual record with the extreme value ranged from 6.5 to 8.25, suggesting that there may be an error in recording, and to be cautious, the value was excluded from the analysis. Although the literature is sparse, there is some evidence examining whether traits measured in foals are the same trait as that measured in a three-year old horse. In the Trakehner, traits measured in foals and mares were found to be very similar, based on sire breeding values estimated from the two data sources (Preisinger *et al.,* 1991). In the Holsteiner, high genetic correlations were found between mare studbook registration data and foal performance tests, although correlations between mare station tests and foal tests were low or moderate (Bösch *et al.,* 2000).

As the dataset is relatively small at present, there were insufficient data to analyse each trait within each discipline. Therefore it was important to assess whether data could be pooled, assuming that the traits as measured in Futurity were genetically identical across disciplines. However, with so few horses evaluated in multiple disciplines, there were insufficient data to estimate genetic correlations across disciplines with sufficiently low standard errors. Correlations between horse effects for the three main disciplines were therefore estimated. The horse effect included all genetic and non-genetic effects, e.g. permanent environment, after having accounted for age, gender and temporary environment. This analysis used only horses with observations in multiple disciplines. High correlations indicate that the genetic correlations are also very likely to be high unless there was very small additive genetic variance for both

traits, which is unlikely. This would provide a justification for the assumption that the traits may be genetically identical across disciplines. Bivariate models including all fixed effects and the random effect of the horse were used to assess correlations. Results (presented below) indicated that the assumption of genetic similarity across disciplines was tenable. Data were therefore pooled over disciplines with any mean differences being accounted for in the models by including discipline as a fixed effect.

4.3.2.2 Genetic analysis

Univariate analyses for each of the five traits were performed using an individual animal model in ASReml (Gilmour *et al.,* 2006).

$$y = mean + dis + loc + sex + loc.date + \beta age + spl(age) + a + c + e$$

The fixed effects were discipline (dis), location (loc), gender (sex) and location – date interaction (loc.date). β is the slope of the regression of the trait on age (age) at the time of the evaluation, however the effect of age was fitted as a spline with the smoothing parameter estimated as part of the random model as described in White *et al.* (1999). With no a-prior knowledge of the shape of the curve relating age to score, splines were used for their flexibility. Random effects were the genetic effect of the horse (a), the permanent environment of the horse (c) and the residual error (e). The random effects a, c and e were assumed normally distributed with (co)variance matrices $\Sigma_a \otimes A$, $\Sigma_c \otimes I$, $\Sigma_e \otimes I$. The significance of random effects was assessed using an approximate t-test, and where the variance component was obviously significantly greater than zero, no further test was done. A likelihood ratio test was performed for those variance components where the significance was less apparent. The significance of fixed effects was examined using an approximate F-test; differences between levels of fixed effects were assessed using the t-test.

Although the data included information on the owner and breeder, the recording of this data was not complete (53.8 % and 53.5 % recorded respectively). These effects were included in preliminary analyses as random effects however they reduced the size of the dataset substantially, and reduced the accuracy of the estimates, and in general, in these preliminary analyses, there was no evidence of effect of breeder/ owner. For this reason, the final model excluded the owner and breeder effects. Multivariate analyses were also performed between all pairs of traits, and the models included the same fixed and random effects as in the univariate analyses.

Functions of variance components were calculated as follows: phenotypic variance, $\sigma^2_p = \sigma^2_a + \sigma^2_c + \sigma^2_e$; heritability, $h^2 = \sigma^2_a / \sigma^2_p$; repeatability, $r^2 = (\sigma^2_a + \sigma^2_c)/ \sigma^2_p$; where, σ^2_a is the additive genetic variance, σ^2_c is the horse permanent environmental variance (i.e. non-genetic) and σ^2_e is the residual error variance for the trait.

4.3.3 Estimated Breeding values

The distribution of estimated breeding values was examined for each trait. Estimated breeding values were scaled by $20/ \sqrt{\sigma^2_a}$, with a mean estimated breeding value of 100. The range and reliabilities of estimated breeding values were also examined, where the reliability was calculated as $1- [se^2 / \sigma^2_a]$, and se is the standard error of the estimated breeding values presented in ASReml.

4.4 Results

4.4.1 Description of dataset

In total there were 1887 evaluation records on 1323 horses (Table 4.1). The number of evaluations increased throughout the years; 218 were recorded in 2006, 278 in 2007, 529 in 2008

and 862 in 2009. There were approximately equal numbers of evaluations for dressage and eventing (35.4% and 34.9% respectively), less for show jumping, sports ponies and endurance horses (19.6%, 8.9% and 1.3% respectively). Due to the small numbers of records for ponies and endurance, results from this point onward are presented only for dressage, eventing and show jumping. The age range for horses at evaluation was 14 days – 1591 days. For horses attending an evaluation date (where it was potentially evaluated for multiple disciplines), 41% were foals, 24% yearlings, 21% 2-year olds, 13% 3-year olds and <1% 4-year olds. Horses often had multiple evaluations in one discipline: for dressage, 10.2% and 1.5% of horses had 2 and 3 evaluations respectively. Figures for show jumping were 9.6 % and 2.2%, and eventing were 15.5% and 1.8%. Horses were also evaluated in multiple disciplines. Of the horses evaluated for dressage (n=587) 14.7% and 18.1% were evaluated for show jumping and eventing respectively, and of the show jumping horses (n=323) 26.6% and 31.0% were evaluated for dressage and eventing.

Table 4.1 Summary of dataset by discipline for all years

Discipline	% of records	% of horses	No. locations	Age (days) Min	Max
dressage	35.4	44.3	14	14	1591
show jumping	19.6	24.3	15	14	1572
eventing	34.9	41.3	15	26	1572
Sport ponies	8.9	10.6	15	34	1527
endurance	1.3	1.9	4	47	1248
Total number	1887	1323			

The percentage of horses evaluated from the total number of horses is given by disipline; as horses were frequently evaluated for multiple disciplines the overall total percentage of horses evaluated is greater than 100.

The total number of sires and dams was 607, and 1096 respectively, which were available for 99.7% and 99.9% of records, and 663 sires of dams were available representing 77.7% of data. The mean progeny size of the half-sib groups was 2.17 (sires) and 1.21 (dams). The maximum sizes of descendant groups were 38, 4 and 11 for sires, dams and dams' sires respectively. Few

of the evaluated horses were also sires (n=6) or dams (n=2) of others that were evaluated, although this is mainly due to the short time period that the data covered. A total of 71 sires were also dams' sires. The genetic connections in the dataset are expected to increase rapidly as data accumulate further.

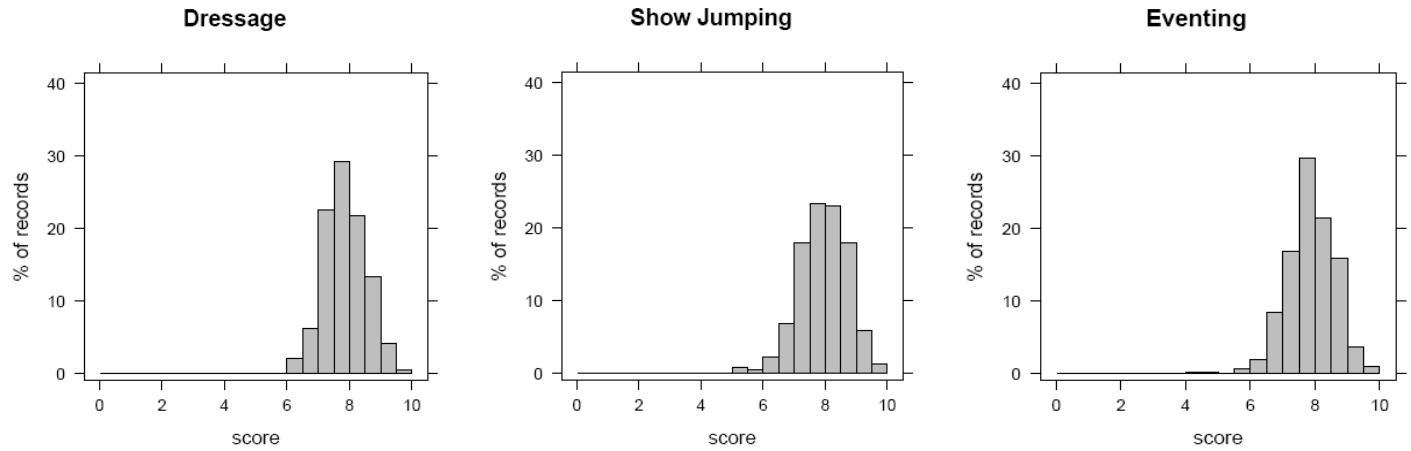
4.4.2 Phenotypic parameters

Summary statistics for the scores for each of the three main disciplines are presented in Table 4.2. The mean score for each trait in each discipline was between 8.02 and 8.24. The standard deviation ranged between 0.54 and 0.83, and was highest for veterinary. The overall minimum in the data analysed was 4 and maximum 10, indicating that the lower range of the 10 point scale is not used, probably due to pre-selection of horses by owners and the guidelines which attempt to define scoring in absolute terms. Distributions were approximately normal with some slight negative skewness, particularly for veterinary. Histograms of the distributions of scores are presented in Figure 4.1.

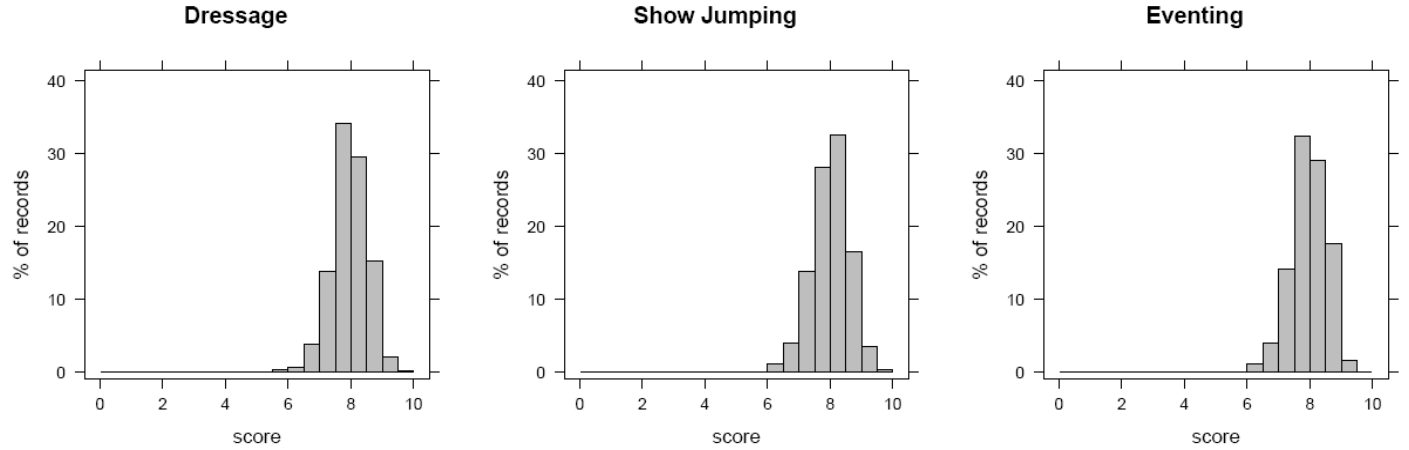
Table 4.2 Summary statistics for variable scores for each trait for the three main disciplines

Trait	Discipline	n	mean	s.d.	min	max
Athleticism	**D**	665	8.02	0.67	6.00	10.00
	SJ	368	8.12	0.78	5.00	10.00
	EV	657	8.05	0.73	4.00	9.75
Conformation	**D**	668	8.09	0.56	5.50	9.75
	SJ	369	8.14	0.61	6.00	9.62
	EV	658	8.11	0.56	6.00	9.50
Correctness of paces	**D**	668	8.08	0.57	4.50	10.00
	SJ	367	8.02	0.76	4.00	9.75
	EV	658	8.05	0.54	5.00	9.75
Type & temperament	**D**	667	8.14	0.60	6.00	10.00
	SJ	367	8.22	0.68	6.00	10.00
	EV	656	8.18	0.63	4.00	10.00
Veterinary	**D**	601	8.14	0.82	4.00	9.75
	SJ	336	8.24	0.81	4.50	9.75
	EV	560	8.18	0.83	4.00	10.00

D=dressage, SJ=showjumping, EV=eventing

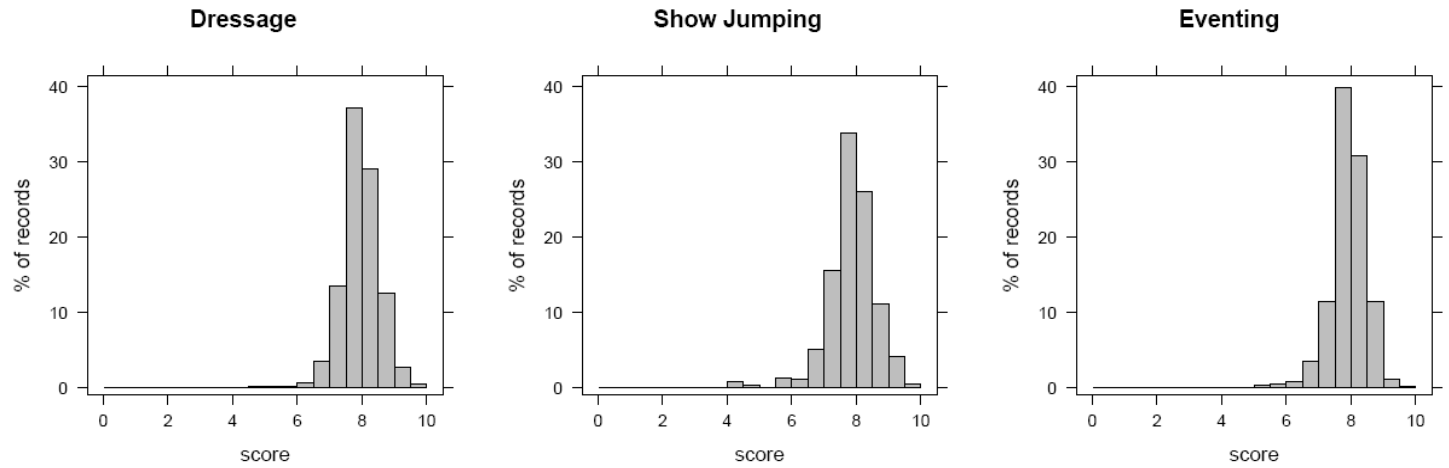


Athleticism

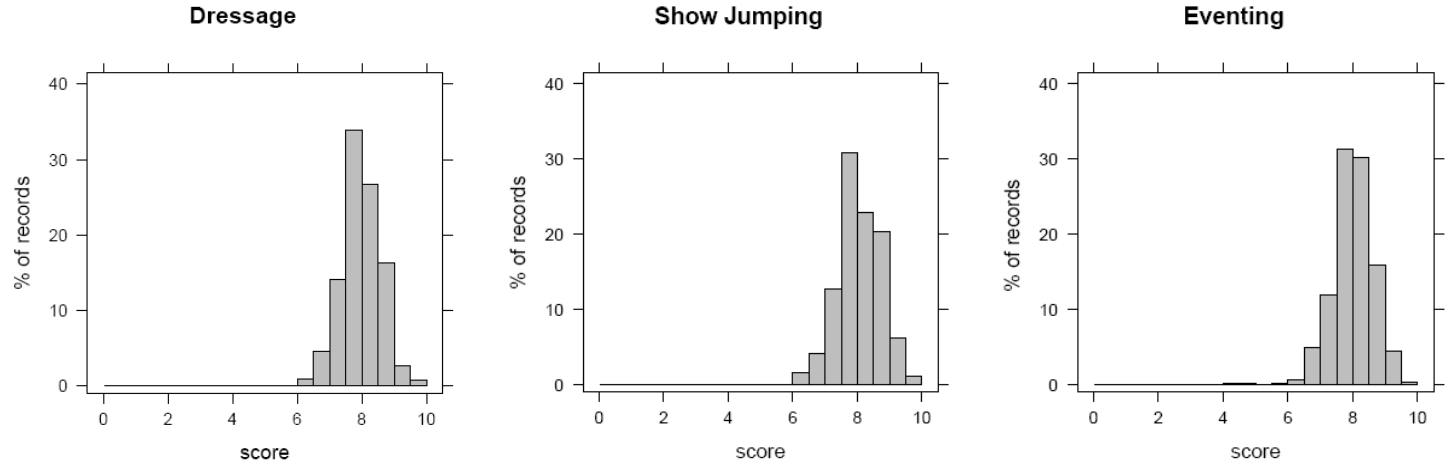


Conformation

Figure 4.1. a) Distributions of scores for each trait, for dressage, showjumping and eventing disciplines for athleticism & conformation

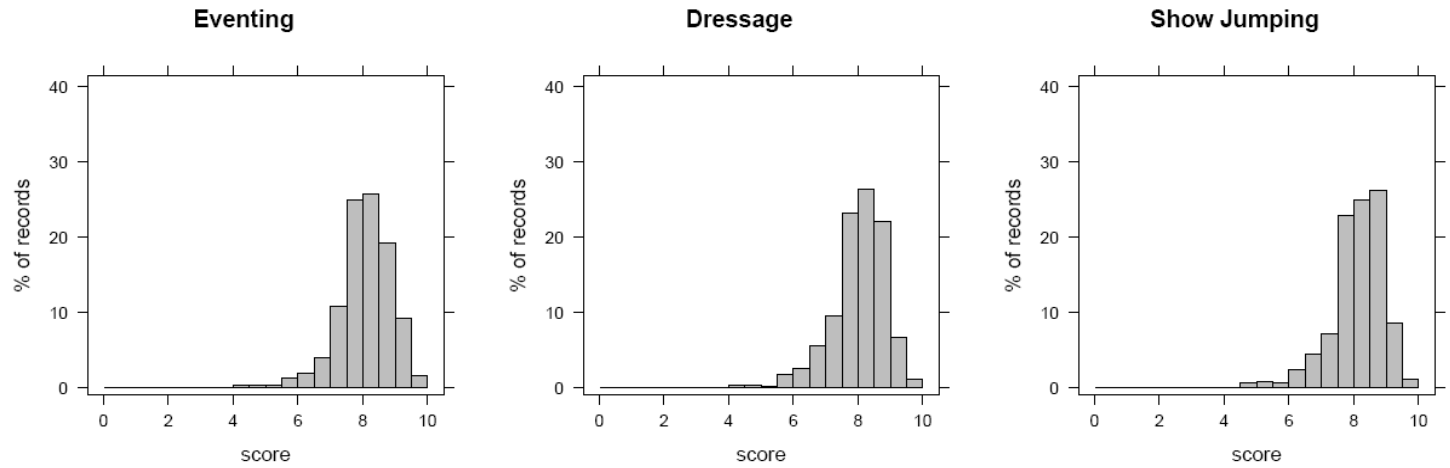


Correctness of paces



Type & temperament

Figure 4.1. b) Distributions of scores for each trait, for dressage, showjumping and eventing disciplines for correctness of paces & type & temperament



Veterinary

Figure 4.1. c) Distributions of scores for each trait, for dressage, showjumping and eventing disciplines for veterinary

In the analysis conducted to examine the feasibility of pooling test data across disciplines, the correlations between the three main disciplines for the random effect of the horse were not significantly different from 1 for any pair of traits (Table 4.3). In the absence of being able to assess genetic correlations between traits directly (due to lack of data), and the magnitude of heritabilities found (reported below), the genetic correlations were also assumed to be high, and it was considered appropriate to pool the data across disciplines.

Table 4.3 Correlations between disciplines for the overall horse effect (additive genetics & permanent environment)

Trait		Discipline	
		SJ	EV
Athleticism	**D**	0.960	0.996
	SJ		0.990
Conformation	**D**	0.990	0.990
	SJ		0.986
Correctness of paces	**D**	0.956	0.975
	SJ		0.960
Type & temperament	**D**	0.971	0.971
	SJ		0.995
Veterinary	**D**	0.996	0.990
	SJ		0.999

D=dressage, SJ=show jumping, EV=eventing
Analyses were based on 86, 104 and 99/98 horses for D-SJ, D-EV and SJ-EV respectively, except for veterinary, where numbers were 82, 98 & 95 respectively.

To consider whether horses that were evaluated for multiple disciplines were a biased sample of the total, within each discipline and trait, the mean score was compared between horses evaluated for multiple disciplines and horses evaluated in one discipline, using a t-test. For dressage, for all traits with the exception of veterinary, the mean score for horses evaluated in multiple disciplines was significantly lower. This was also true for show jumping for type and temperament and athleticism. This indicates that there is some evidence for show jumping and

dressage that either horses evaluated just one discipline are better horses, or that they have received more training for that discipline. For show jumping the difference between groups was apparent for traits that are improved by training, suggesting the latter may be true. There were no differences between groups for cross-country, which again may reflect less specialised training for this discipline at such a young age.

4.4.3 Genetic parameters for single traits

Heritabilities, repeatabilities and the proportion of variance due to the horse's permanent environment for each trait are shown in Table 4.4. All were significant ($P<0.05$) with the exception of permanent environment for type and temperament. Heritabilities ranged from 20.34% to 42.16%, and were highest for type and temperament, and lowest for athleticism. The ratio of the variance due to the horse's permanent environment to the total phenotypic variance ranged between approximately 2.80% (not significant $P>0.05$) for type and temperament and 51.64% for veterinary. The repeatability, representing the total variance due to the horse, including both additive genetic and permanent environmental variance, was lowest for type and temperament and athleticism and highest for veterinary. The residual error – unexplained variance in scoring - may include the effects of the handler, owner and breeder, although some of the longer term influences of the latter may also be in the permanent environment.

Table 4.4 Heritability, repeatability and horse's permanent environment as a percentage of total phenotypic variance, for each trait, with standard errors in parentheses

Trait	Additive genetics	Horse's permanent environment	Residual error	Repeatability	Total phenotypic variance
Athleticism	20.34 (8.28)*	25.63 (8.58)**	54.03 (3.39)***	45.97 (3.39)***	0.46
Conformation	29.05 (9.36)**	35.09 (9.38)***	35.86 (2.46)***	64.14 (2.46)***	0.30
Correctness of paces	30.42 (9.28)**	24.63 (9.32)**	44.95 (2.95)***	55.05 (2.95)***	0.33
Type & temperament	42.16 (9.70)***	2.80 (9.49)	55.04 (3.47)***	44.96 (3.47)***	0.37
Veterinary	24.95 (11.01)*	51.64 (10.99)***	23.41 (1.82)***	76.59 (1.82)***	0.48

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

4.4.4 Fixed effects of gender, discipline and age

Gender effects are given in Table 4.5. In general geldings scored lower than stallions, although differences were only significant for veterinary and athleticism ($P<0.01$). These differences are partly due to the fact that the best males are kept as stallions. The mean score awarded also differed between disciplines (Table 4.6). In general, the traits were scored more harshly for dressage: for instance, in type and temperament, show jumping and eventing scored higher than dressage ($P<0.01$). Show jumping was scored lower than eventing for correctness of paces ($P<0.01$), but there were no other significant differences between the two disciplines.

Table 4.5 Effect of gender on scores awarded.

Trait	Gender	Stallion (0 comparison)	Gelding (0 comparison)
Athleticism	**stallion**		
	gelding	-0.13**	
	mare	-0.02	0.11**
Conformation	**stallion**		
	gelding	0.01	
	mare	0.04	0.03
Correctness of paces	**stallion**		
	gelding	-0.08	
	mare	0.02	0.10**
Type & temperament	**stallion**		
	gelding	-0.08	
	mare	-0.02	0.06
Veterinary	**stallion**		
	gelding	-0.18**	
	mare	-0.11**	0.06

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

Table 4.6 Effect of discipline on scores awarded.

Trait		Discipline D (0 comparison)	SJ (0 comparison)
Athleticism	**SJ**	0.10*	
	EV	0.13***	0.03
Conformation	**SJ**	0.04	
	EV	0.07*	0.03
Correctness of paces	**SJ**	-0.04	
	EV	0.07*	0.10**
Type & temperament	**SJ**	0.11**	
	EV	0.14***	0.02
Veterinary	**SJ**	0.04	
	EV	0.01	-0.03

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$
D=dressage, SJ=show jumping, EV=eventing

Results for the effect of age on the scores for each of the traits are shown in Figure 4.2. In general, with the exception of veterinary scoring, the youngest horses performed relatively well. Scores then declined with age until the horse was a yearling, and then rose again as the horse matured. This implies that scores tend to increase over age, as the horse matures and so temperament, paces etc. improve. However very young horses are either difficult to assess, or tend to be given higher scores through subjective bias towards foals. For veterinary, the foals were scored lowest, and overall scores increased with age. For all traits, high scores were seen in three year olds. Such a trend may reflect improvement of traits, but there are other explanations including selection in submitting for an evaluation. Judges may also be better at discriminating traits in more mature horses. As horses may have been evaluated on several occasions, any improvement in scoring due to experience would give a false indication of an age related change. This is particularly relevant to type and temperament where experience is very likely to improve scoring.

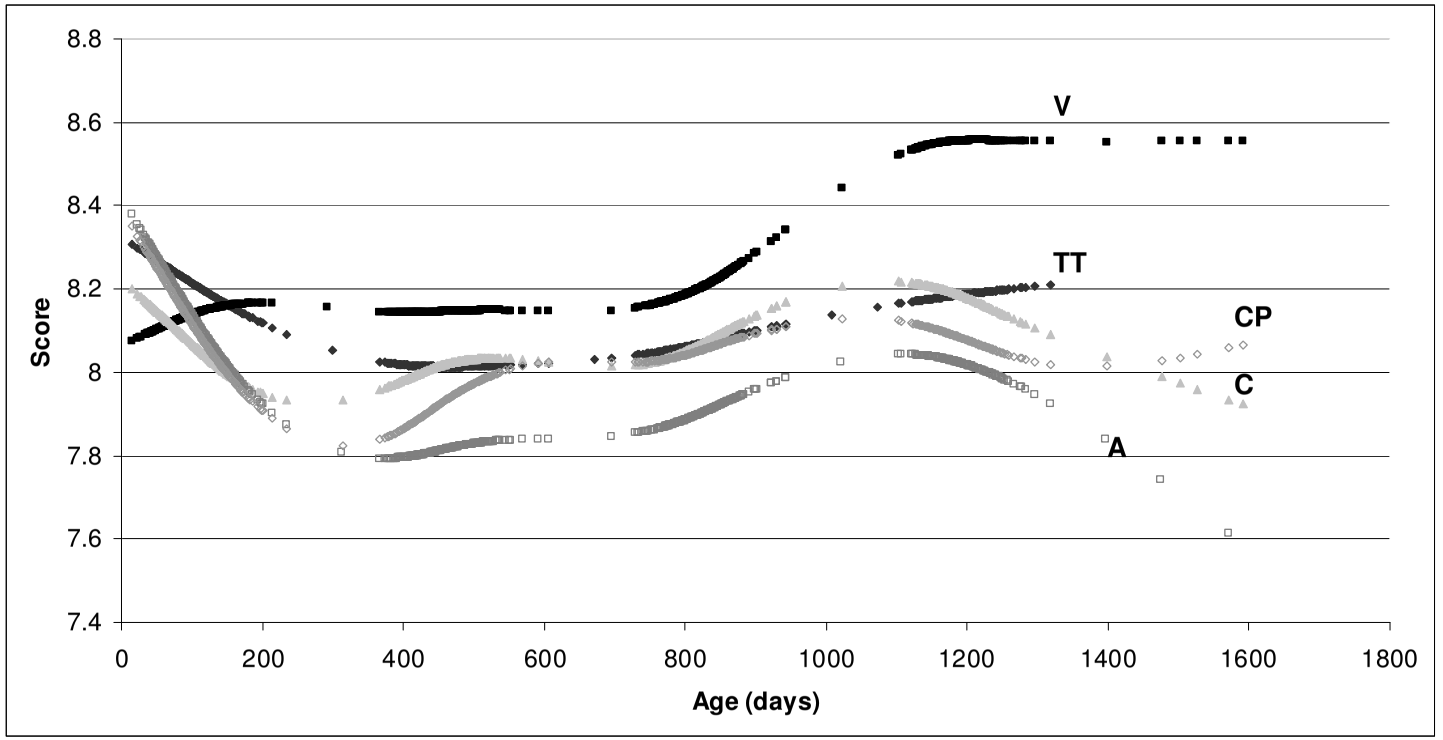


Figure 4.2. Influence of age on scoring for each trait (athleticism (A) = open medium grey squares, conformation (C) = solid light grey triangles, correctness of paces (CP) = open medium-light grey diamonds, type and temperament (TT) = solid dark grey diamonds, veterinary (V) = solid black squares)

4.4.5 Genetic correlations between traits

The genetic correlations between traits were generally high and all were significantly different from zero, where standard errors were determined ($P<0.05$) (standard errors were not determined for athleticism - type and temperament, athleticism - correctness of paces and type and temperament - correctness of paces). Apart from correctness of paces and veterinary, and correctness of paces and conformation, they were not significantly different from 1.0 (Table 4.7.). Correlations between traits other than veterinary were particularly high. The magnitude of these correlations indicates that the same genes are largely responsible for all traits.

Table 4.7 Genetic correlations (standard error) between traits

Trait	Conformation	Correctness of Paces	Type & Temperament	Veterinary
Athleticism	0.94 (0.09)	1.00 [a]	1.00[a]	0.57 (0.23)
Conformation		0.77 (0.11)**	0.97 (0.05)	0.83 (0.15)
Correctness of Paces			1.00 [a]	0.44 (0.22)*
Type & Temperament				0.80 (0.17)

Significant difference from 1: * = $P<0.05$, ** = $P<0.01$
[a] standard errors could not be determined in the analysis

4.4.6 Permanent environment correlations between traits

Correlations between traits for the effect of the permanent environment were significant ($P<0.05$) and moderate between correctness of paces and all other traits, and the correlation between type and temperament and athleticism was moderate and significant (Table 4.8.). No significant correlations were detected between the other pairs of traits. Thus, compared to the genetic correlations they were smaller and not all significantly different from zero. This indicates that the same aspects of the horses' permanent environment, such as early life influences, stabling, and long-term training and long-term nutritional effects, jointly influence some of the

pairs of traits. Also included in the permanent environment effect in this analysis will be maternal effects and non-additive genetic effects.

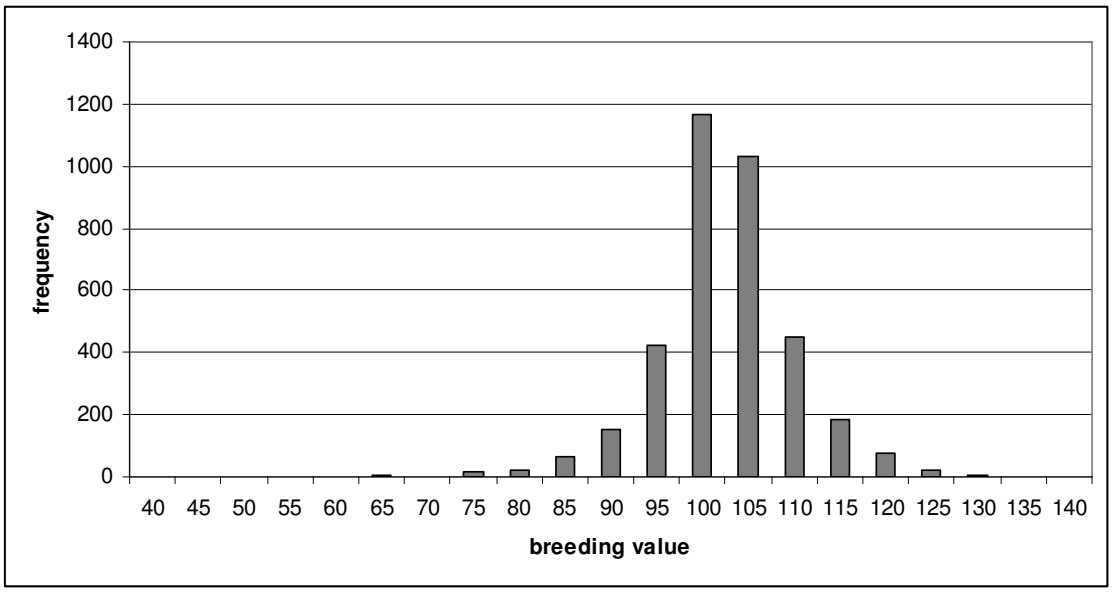
Table 4.8. Permanent environment correlations (standard error) between traits

Trait	Conformation	Correctness of Paces	Type & Temperament	Veterinary
Athleticism	0.37 (0.19)	0.69 (0.05)***	0.58 (0.12)***	0.24 (0.18)
Conformation		0.65 (0.13)***	0.60 (0.34)	0.24 (0.16)
Correctness of Paces			0.46 (0.13)***	0.43 (0.18)*
Type & Temperament				-0.02 (0.59)

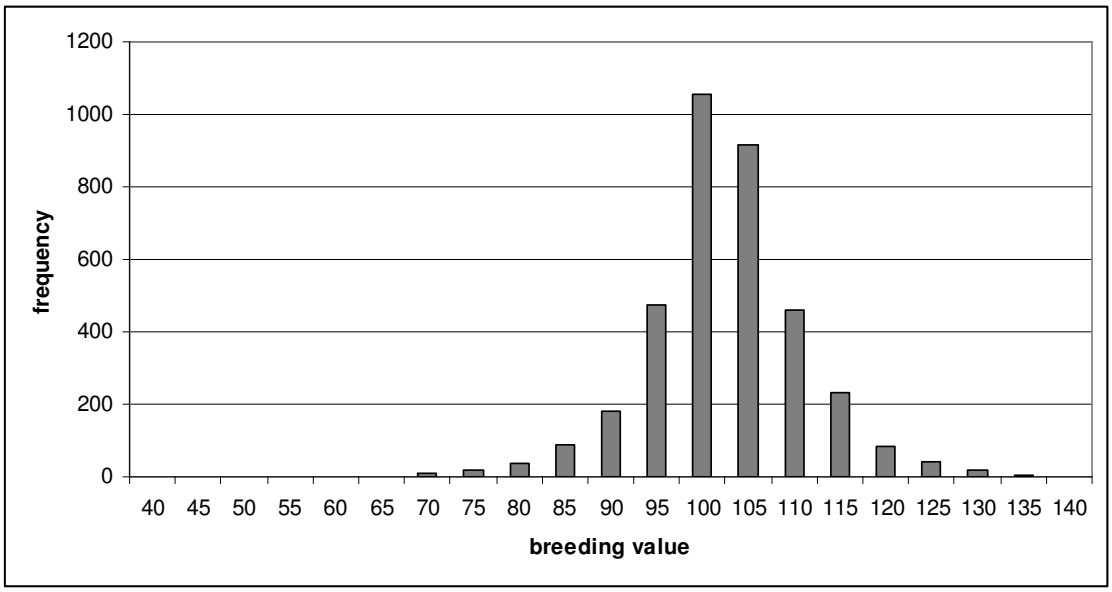
Significant difference from 1: * = $P<0.05$, ** = $P<0.01$, *** = $P<0.01$

4.4.7 Estimated Breeding values

The distribution of estimated breeding values, for all animals in the pedigree, for the examples of athleticism and conformation are given in Figure 4.3. The highest reliabilities for estimated breeding values were 0.81 (type and temperament), 0.74 (correctness of paces), 0.72 (conformation), and 0.66 (veterinary and athleticism). For publication, industry has decided that only estimated breeding values with a minimum reliability of 30% will be produced. In these data 39.3% of horses would quality for type and temperament, 31.9% for correctness of paces, 37.5% for conformation, 10.7% for veterinary and 8.4% for athleticism.



Athleticism



Conformation

Figure 4.3. Distribution of estimated breeding values for athleticism and conformation

4.5 Discussion

The young horse tests (Futurity) recently introduced in the UK are a valuable data source for genetic evaluations. Participation has grown substantially since the introduction of the scheme, and it is envisaged that the data will accumulate rapidly. The most appropriate strategy will then be to combine young horse data with adult competition data to routinely estimate breeding values. This study characterised the Futurity dataset to date and performed the first genetic evaluations on traits recorded on young horses in the UK.

To date, genetic evaluations in the UK have been conducted as a research exercise for dressage and eventing using competition data (Stewart *et al.,* 2010; Kearsley *et al.,* 2008), as large competition datasets exist. In general, both young horse tests and adult competition data have advantages and caveats, and now most international evaluations use a combination of both data sources. Competition data may be available on a greater proportion of the population however it becomes available late in life and is generally biased in that the horse may compete in only one discipline. Young horse data are obtained much earlier in life, which enables a shorter generation interval thus providing opportunity for greater genetic gain. To be useful, young horse data generally require moderate to high heritabilities for the traits, and high genetic correlations with later competition performance.

Selection in the horses evaluated at young horse tests is present in most evaluations internationally. The effect of this will be to bias heritability estimates downwards. In the UK, an estimated 10% of the 2009 sport horse foal crop was evaluated in the Futurity scheme of that year (Graham Suggett, personal communication). The actual proportion of young horses evaluated at any age is therefore at least 10% and possibly at much as 30%. This is in line with

the percentages of young horses tested internationally, which range from 10% to 45% of the (registered) foals (Thorén Hellsten *et al.,* 2006). The Futurity is therefore capturing a comparable proportion of young horses to other international young horse tests, while only in its third year and participation is still growing. Selection in the horses presented is illustrated in the lack of scores in the lower range of the scale. There is anecdotal evidence that the degree of selection has increased over the years (J. Rogers personal communication), as owners become more aware of the characteristics required. A higher proportion of youngstock from stallions with high stud fees tend to be entered (J. Rogers personal communication), and from these stud farms, it is likely that only a limited number of foals are evaluated, suggesting that selection is of the gene pool identified as being more elite. There may be other biases in the horses that are evaluated, for instance related to the fact that participation tends to come from the same owners, and there are limitations in presenting horses for evaluation including the cost of evaluation, and practical resources such as travel costs involved in transporting a horse to the venue.

A team of judges evaluate the traits. In the current analysis, the variance due to the judging teams will be included with the temporary environmental variance (location–date interaction). Training prior to evaluations ensures consistency and teams vary as individual judges rotate between different teams (based on practical issues rather than random assignment or specific groupings). In the years covering the data collection, scoring was done by discussion within the team, which may have led to problems, for instance a more dominant judge may exert an undue influence on the scoring. This strong influence of the judging team has been shown by Preisinger *et al*., (1991). As of 2011, scoring will be performed by judges individually, and then averaged, which will allow the repeatability of scoring to be evaluated.

The UK tests are distinct in a number of ways from any others internationally. One difference is the ages being evaluated - UK tests cover foals up to 3 (4) year olds, whereas international tests range from 3 to 7 year olds (Thorén Hellsten *et al.,* 2006). UK tests also evaluate for only one discipline, although disciplines include eventing, endurance and ponies, whereas international tests often evaluate one horse for both show jumping and dressage. Internationally, eventing evaluations are uncommon. The duration of testing is also relatively short compared to other national tests. UK tests are most comparable to field tests, rather than the station tests or young horse competitions that take place in some other countries. Station tests generally have higher heritabilities compared to field tests, at least in part due to the control of environmental factors (Ricard *et al.,* 2000). The longer length of station tests may also produce higher heritability estimates (Dietl *et al.,* 2005). Due to the age constraints there is no testing under saddle in the UK whereas tests elsewhere are either ridden or at hand; tests performed without the rider have higher heritabilities than those with the rider, due to the lack of rider variance (Ricard *et al.,* 2000). At present, all genders are eligible to participate, without restrictions. However, as of 2012, the eligibility of stallions may change, so that only graded stallions (with studbooks) are eligible, similar to practices on the continent. This change to the rules will introduce bias to the genetic evaluation system, producing downwardly biased heritability estimates due to the restricted genetic pool and thus lower genetic variance. In many foreign evaluations, the breeding goals include subjective terms (Koenen & Aldridge, 2002). This is also true of traits assessed in the Futurity, illustrated for example, by the descriptions of “confidence”, “expression”, “harmony” and “honest” within the description of traits. Such subjectivity may result in an increase in phenotypic variance and a reduction in both repeatability and heritability estimates.

This analysis was performed on a relatively small dataset, due to which we had to address strategies of analysing small datasets. Phenotypic correlations between disciplines were estimated. This indicated that genetic correlations were consistent with unity, and justified to an extent the pooling of traits across disciplines in the analysis. The assumption of equal traits across disciplines should be further confirmed when more data are available (however, as horses are now evaluated for only one discipline at any one event, suitable data may take time to accumulate). In contrast to these high estimates, current foreign estimates for genetic correlations between dressage and show jumping related traits from young horse data are mixed (low and either positive or negative), although correlations between canter and jumping tend to be higher than correlations between the other gaits and jumping (Thorén Hellsten *et al.,* 2006). If the genetic correlations are indeed higher in the UK then this may well be due to the lower age of testing, when differentiation between disciplines has not yet occurred. There may be various advantages and disadvantages to testing at such a young age. The influence of selection due to training and competing in young horse competitions is reduced, and the earlier the data become available for genetic evaluation, the greater the potential for genetic progress. However, it is more difficult to assess future ability in very young horses, which are still developing rapidly.

Perhaps the most comparable test internationally is that of the Swedish Young Horse Tests (YHT) for 3 and 4 year olds. In these tests, the heritability of total conformation was 0.58, which is twice the estimate of 0.29 from the current study. The heritability of type in the Swedish tests was 0.30, which compares with a higher estimate of 0.42 in the current study. Heritabilities for walk and trot at hand and for free canter were 0.37, 0.45 and 0.37 respectively, i.e. slightly higher than the 0.30 for correctness of paces in this study (Viklund *et al.,* 2008). In other countries overall, the average estimates from field performance tests were 0.30, 0.35 and 0.28 for walk, trot and canter, i.e. very close to the 0.30 estimate for correctness of paces in the

current study (Thorén Hellsten *et al.,* 2006). Thus although there are some differences between estimates in the literature and the current study, other estimates are very close and are of a similar magnitude overall.

The heritability was highest for type and temperament. Type, reflecting physical characteristics, might be expected to be very influenced by genetics. Temperament might be expected to be influenced by both temporary and permanent environment, including early life influences and training. However, no significant variance due to the horse's permanent environment was detected for type and temperament. Heritability estimates for temperament are scarce in the literature. Oki *et al.* (2007) estimated the heritability of temperament in the thoroughbred as 0.23. For young horse tests in Dutch warmbloods, heritability of character was estimated at 0.52 for station tests and 0.06 for field tests (Huizinga *et al.,* 1991 & Huizinga *et al.,* 1990). In the RHQT (1988 – 2003) of the Swedish warmblood, heritabilities of temperament for jumping and gaits were 0.17 and 0.41 respectively, and for the YHT, temperament for jumping was 0.23 (Viklund *et al.,* 2008).

For use in genetic evaluations, young horse trait data must show moderate/ high heritabilities and high genetic correlations with later competition performance. The next requirement in these studies is to investigate the genetic correlations between the young horse data and adult competition data. Internationally, estimates for genetic correlations between young horse tests and adult competition are generally positive and moderate/ high indicating that the young horse tests are effective in selecting for adult competition performance (Thorén Hellsten *et al.,* 2006). For example Viklund *et al.* (2010) estimated correlations between competition data and traits recorded in the Swedish RHQT, and found the correlation between show jumping competition

success and the jumping traits was 0.87 to 0.89 (depending on the competition performance trait) and between dressage competition success and the average for gaits was 0.68 to 0.75.

Pedigree data/ recording are often a major limitation in the genetic evaluation of horses. The pedigree available in this analysis was shallow (including only sire, dam and dam's sire). A greater depth of pedigree, which should accumulate over time, would provide more accurate estimates of the genetic variance. The young horse data were recorded comparatively well, however horses are identified by name only. The use of a unique identifier, such as the Universal Equine Life Number or passport number would greatly add to the integrity of the data. The recording of the pedigree data is of particular importance for the purposes of genetic evaluations and this is an issue that will escalate over time unless resolved. This analysis was performed on the first four years of data and was therefore based on a relatively small dataset. As the data accumulate over the years, the accuracy of estimates and thus extent of interpretations that can be drawn will increase.

4.6 Conclusion

The young horse tests introduced in GB have been successful and participation continues to grow. Equal numbers of evaluations are for dressage and cross-country, with less in show jumping. The data are a potentially valuable data source for genetic evaluations. The magnitudes of heritabilities indicate that, if genetic correlations between scored traits and later competition performance are high, the data will be useful in selecting for later competition performance. At present, the reliability of estimated breeding values is low to moderate but reliabilities will increase rapidly with additional data. The young horse data could be used to predict breeding values for the traits recorded as part of the scheme although best use of the data would be made

by combining it with adult competition data for use in the prediction of breeding values for competition.

5. Evaluation of British Horses Competing in International Showjumping Competitions

5.1 Abstract

Genetic evaluations for showjumping are performed by many Northern European studbooks, however no evaluations have yet been performed in the UK. The aim of this work was to perform genetic evaluations for showjumping in GB, using results of British competitors competing in elite national and International competitions.

Competition data from 2007 – 2009 was obtained from British Showjumping. Performance measures used were the normal score of place and the log of penalty points. Univariate analyses were performed in ASReml. Fixed effects were the height of the jump, horse age and show, and random effects were the combined genetic and permanent environment effect of the horse, the competition group nested within show and the residual error. An upper limit on the heritability was estimated. Further models attempted to examine the effect of studbook, the coursebuilder, indoor/ outdoor competition, the proportion of horses with time faults and the ground conditions.

There were a total of 6239 records in the model using the normal score of place as a performance measure. In this model, an upper limit on the heritability was estimated as 8.9%, and the variance due to class within show accounted for 2.6% of the total phenotypic variance. Age and height significantly affected performance, and in general the levels of studbook did not have a significant effect on performance.

The degree of selection in the competition data was very high and limited the validity of the variance estimates derived. However, results indicate that there is significant genetic variance in the population for performance in showjumping, and this value is likely to be higher than the estimate from this study. To perform more accurate genetic evaluations, enabling selection for showjumping performance, it is necessary that more substantial data sources are available.

5.2 Introduction

One of the main breeding objectives for sport horses in GB, and internationally, is improved performance in showjumping. Many of the studbooks that are most successful in producing horses which excel in high levels of showjumping competition perform genetic evaluations and produce EBVs. For instance, evaluations are performed by the national studbooks of Sweden, Ireland, France, Belgium and Denmark, Germany and Holland (Janssens, 2008; Ruhlmann *et al*., 2009a). EBVs aid in the selection of animals for breeding, which in turn achieves genetic progress for the sport.

Genetic evaluations frequently use adult competition data, often in association with young horse test data. Competition data has been recorded by studbooks for long time periods, and so large datasets are available. The performance traits used differ, and include either individual competition records, or an aggregate annual or lifetime measure, for instance rank (Tavernier, 1990), normalised ranks (Reilly *et al*., 1998), annual earnings (Langlois, 1980; Ricard & Chanu, 2001), lifetime cumulative points and placings (Wallin *et al*., 2003). These may be transformed to achieve more normal distributions of the data, suitable for analysis. GB does not currently perform genetic evaluations, or produce estimated breeding values, which puts the GB breeding program at a major disadvantage compared to foreign competitors.

Heritability estimates for performance include 0.08/ 0.05/ 0.09 (rank for low/medium/high levels of competition) for the Irish sport horse (Ruhlmann *et al*., 2008), 0.10 for the Hanoverian, Trakehner, Oldenburg and Danish Warmblood), 0.15 for KWPN, 0.16 (ranks) and 0.27 (log(annual earnings)) for the Selle Français (Janssens, 2008), and 0.27 (cumulative lifetime points) in the Swedish Warmblood (Olsson *et al*., 2008). Predictably, higher estimates are generated from aggregate or cumulative measures. The magnitudes of the estimates indicate that genetic progress can be made in the population through selection. Correspondingly, there is evidence for this occurring in some populations in the past, for instance a genetic gain of 0.93 genetic standard deviations was estimated for showjumping since 1980 in the Swedish warmblood (Thorén Hellsten *et al*., 2009b), and 0.096 annually since 1995 in the SF (Dubois & Ricard, 2007).

British Showjumping (BS), the national competition regulatory body in GB records competition records in two forms: i) core information on all competitions for qualifying classes where applicable, and ii) extended information on international competitions and some national classes. This latter section has been recorded from 2007 onwards. The latter was recorded with the aim of investigating specific aspects of the horse's performance. For instance specific questions to be addressed included:

i) how many horses compete by age group and level

ii) identify horses that perform consistently well by jumping double clears

iii) identify horses that are inconsistent in performance

iv) examine patterns in performance information - trends or traits across age groups

v) identify horses that perform well but have low prize money

vi) identify horses that perform inconsistently but have high prize money

vii) identify horses that have time faults regularly

viii) identify the number of clear rounds jumped at 1.45m and above by age and year.

The extended data was made available for this study for genetic evaluations.

Internationally, most genetic evaluations are conducted by studbooks, and therefore refer to a relatively homogeneous gene pool. Therefore, the genetic diversity is limited. Although there is now much mixing between warmblood breeds, the extent of mixing has always to some extent been limited by the regulations of studbooks, and today, breeds have become such a mixture that the distinction between them is less and there is greater homogeneity between breeds. Only the German national evaluation includes a breed effect in the evaluation model, but it already refers to a relatively homogeneous gene pool of German warmbloods. Recent work by Thorén Hellsten *et al*. (2009b) considered the inclusion of the effect of breed in the Swedish national evaluation. However, it was deemed to be inappropriate, due to the fact that there are such large pedigrees available and that the Swedish warmblood is now such a mixture of warmblood breeds. The GB data however, is recorded by the competition body, as opposed to a studbook. It therefore may refer to a wider range of breeds. At elite competition levels, the range of breeds may be somewhat restricted, nevertheless there will be a variety of breeds represented, and it may be of interest to look at the effect of breed in the GB data.

Three years of extended competition data, covering international and selected national classes, are now available for analyses. The aims of this work were therefore to i) examine the suitability of this data for genetic analysis, ii) to estimate the genetic parameters needed for breeding value prediction and to predict breeding values for all horses in the pedigree.

5.3 Materials and Methods

5.3.1 Data

BS supplied competition data covering international and certain national classes of competition data, referring to horses/riders registered with the BS. Three years of data were represented: 2007 – 2009. In 2008, the data was streamlined due to resource restrictions and the immediate requirements of the BS, and in 2009 recording was further streamlined, to include only results for horses that were consistently performing well (scoring double clears at 1.45m and above).

Variables relating to the rider, horse, competition and show were available. Rider variables included age and Fédération Equestre Internationale (FEI) ranking. Horse variables included age, BS identity, and breed. Gender data was obtained from the BEF and from the BS. Competition data included number of starters, height of jumps, class type and class number. Show data included show and date. In addition, in line with the initial aims of the BS, more thorough data on the course, including: the number of riders competing in the show that were within the top 100 ranking by the FEI, course builder, course plan, number of clear rounds, number of horses with time faults, ground conditions, and number of starts. However, in general these were recorded on only limited numbers of records. Pedigree data comprising the sire, dam and dam's sire was available.

5.3.2 Data editing/ manipulation

Records where the horse had abandoned the competition, been eliminated, did not start or retired were excluded from the analysis. In addition, when analysing penalty points as the performance measure, some classes were excluded, namely: top score, accumulator, puissance and 6 bar (n records = 413). Equal placings within a group of competing horses were made ties. This was

based on the number of horses within the data that had that place (i.e. considers British competitors, but not other foreign competitors), and thus was an approximation.

A group of horses competing against each other was defined using the show, date, class type and class number, height of jumps and number of starters. Where the heights of the jumps were given as a range, the mean value was used.

5.3.3 Analysis

Performance measures available were the place obtained in the competition, and the score (i.e. time and jumping penalty points and the time taken). The final penalty point score was calculated as a sum of the time and jumping penalties.

The place was transformed to a normal score, using the number of starters supplied in the data. This follows the introduction of the use of normal scores for genetic evaluations of sport horses by Reilly *et al*. (1998), and subsequent use by others including Kearsley *et al*. (2008). The distribution of the penalty points was examined. The data was right skew, with a large proportion of scores with zero penalty points. To achieve a more normal distribution, the data was transformed using a natural log.

5.3.4 Genetic analysis

Univariate analyses for the normal score of place (Model 1) and the natural log of the penalty points (Model 2) were performed in ASReml (Gilmour *et al*., 2006), using the following model:

$$y = \text{mean} + \text{height} + \text{age} + \text{age}^2 + \text{show} + h + \text{show.cg} + e$$

The fixed effects were height of the jump (height), age (linear and quadratic; fitted based on a preliminary analysis) and show code (show). Random effects were the combined genetic and permanent environment effect of the horse (h), the competition group within show (show.cg) and the residual error (e). The random effects h, show.cg and e were assumed normally distributed with (co)variance matrices $\Sigma_h \otimes I$, $\Sigma_{show.cg} \otimes I$ and $\Sigma_e \otimes I$. The significance of random effects was examined using an approximate t-test, and where the variance component was obviously significantly greater than zero, no further test was done. A likelihood ratio test was performed for those variance components where the significance was less apparent. The significance of fixed effects was examined using an approximate F-test; differences between levels of fixed effects were assessed using the t-test.

Initially, an individual animal model, and sire model were investigated. However, limitations in the size of the dataset prevented the separate estimation of variance components due to horse genetics and permanent environment. Models were therefore adapted, to include the combined effect of the horse's additive genetics and the permanent environment, which will give an upper limit to the heritability. The gender of the horse and the grade of competition were considered as fixed effects in initial analyses, however these were found not to be significantly associated with the trait. Rider was considered as a random effect, however, as one horse is usually ridden by one rider throughout the data (although one rider may ride more than one horse) the random effects of the rider and horse may be confounded, so rider was removed from the final model (although the random effect of rider was found to explain a significant proportion of variance, assessed using an LRT test).

Studbook was fitted as a fixed effect in further models (Models 3 for place and 4 for penalty points). This however excluded 932 and 912 records in models 1 and 2 respectively, due to

missing data. The recording of additional variables relating to the course - such as the course builder and indoors/ outdoors was incomplete. A break-down of the numbers of records are given in Table 5.1. The vast majority of these variables could not be included in the genetic models, due to a lack of data. However, the recording of the number of clear rounds and the number of horses with time faults was high (for 92% of data). These are likely to reflect the difficulty of the course. In further Models 5/6 (place) and 7/8 (penalty points), these were included as fixed effects.

In a further model (Model 9) the proportion of horses that completed clear rounds, and the ground conditions (a categorical variable grass outdoors, outdoors, sand indoors, sand outdoors) were fitted as fixed effects in model 1, this was for 2008 data only.

In summary, the models used were:

Models 1 (place) and 2 (penalty points) as explained above

Models 3 (place) and 4 (penalty points) included studbook as a fixed effect

Models 5 (course builder) and 6 (indoors/ outdoors) (place)

Models 7 (course builder) and 8 (indoors/ outdoors) (penalty points)

Model 9 model 1 plus ground conditions and number of horses with clear rounds (2008 data only).

A likelihood ratio test was used to assess the significance of random effects, where the difference in likelihood was compared to the critical value of the χ^2 distribution with one degree of freedom.

Table 5.1. Variables available from the British Showjumping association

	Number of records with variable recorded:							
Year	**Total no. records**	**No. riders @ show in FEI Top 100**	**Course-builder**	**Course Plan**	**No. Clear Rounds**	**No. horses with time faults**	**Ground Conditions**	**No. starts**
2007	6649	22	0	128	5937	5896	0	801
2008	2989	0	0	2989	2858	2858	2804	0
2009	1456	0	0	0	1453	1453	0	0
Total	**11094**	**22**	**0**	**3117**	**10248**	**10207**	**2804**	**801**

5.3.5 Functions of variance components

Functions of variance components were calculated as follows: phenotypic variance, $\sigma^2_p = \sigma^2_h + \sigma^2_{show.cg} + \sigma^2_e$; upper limit on heritability = repeatability = σ^2_h / σ^2_p; where,σ^2_h is the total horse (additive genetic plus permanent environment variance), $\sigma^2_{show.cg}$ is the variance of competition group within show and σ^2_e is the residual error variance for the trait.

5.4 Results

5.4.1 Description of dataset

A summary of the datasets analysed in the two models with different performance measures is presented in Table 5.2. The distribution of place and penalty points are shown in Figure 5.1 (a and b). In total there were 6239 and 6395 competition records in Model 1 and Model 2 respectively. As datasets were fairly similar, the dataset for Model 2 is discussed further here. Across competition years, the records were distributed as follows: 69.9%, in 2007, 20.1% in 2008 and 10.0% in 2009. The number of records decreased throughout the years, for reasons described previously.

The records covered 475 horses, competing with 156 riders. Generally, one horse was exclusive to one rider; however, one rider rode multiple horses. The average number of records per horse was 13.5, and the number of horses per rider was 3.5, giving an average 41.0 records per rider. The mean horse age at competition was 9.6 years. Horses competed in 3 grades, although 52% were in grade A, and 22% of data was missing. 50.2% of the records were for geldings, 31.8% for mares, and 16.4% for stallions. 455 shows were represented over the three years, with an average of 14 records per show (range 1 – 195). There were 2141 competing groups, with an average of 3.0 records in the data relating to British competitors in each competing group,

although the total number of starters in each competing group overall (including foreign competitors) was much greater, on average 58.7 (range 9 – 194). The height of the jumps was, on average 142 cm (range 110 – 160 cm).

Studbook data was available for 5307 and 5483 records respectively for Models 1 and 2. Horses were from 29 different studbooks. The highest percentages of records were for KWPN (approx 34%) followed by Belgian Warmblood (approx 12%) and then Holsteiner (approx 11%).

Table 5.2.a Summary of datasets analysed in models using normal score of place and penalty points

Variable	Statistic	Model	
		Normal score of place	ln (Penalty points)
Records	**N**	6239	6395
Riders	**N**	155	156
Horses	**N**	479	475
Age (at competiton)	**Mean**	9.43	9.56
	Min	3	3
	Max	17	17
	s.d.	2.67	2.67
Gender			
G	**N**	3110	3213
M	**N**	2014	2032
S	**N**	1014	1047
Missing	**N**	101	103
Year			
2007	**N**	4877	4473
2008	**N**	1362	1284
2009	**N**	0	638
show	**N**	296	455
number records per show	**Mean**	21.08	14.05
	Min	1	1
	Max	200	195
	s.d.	32.92	26.10
Competing groups	**N**	2025	2141
number records within	**Mean**	3.08	2.99
dataset per group	**Min**	1	1
	Max	21	21
	s.d.	2.80	2.70
Height of jump	**Mean**	141.36	142.07
	Min	110	110
	Max	160	160
	s.d.	10.10	10.18
Number of starters	**Average**	80.34	58.71
	Min	7	9
	Max	194	194
	s.d.	45.30	31.35
sires (that are competing horses)	**N**	247 (4)	243 (4)
dams (that are competing horses)	**N**	324 (0)	321(0)

Table 5.2.b Summary of datasets analysed in models using normal score of place and penalty points

		Model	
Studbook	**Breed**	**Normal score of place**	**ln (Penalty points)**
Missing		932	912
Studbook Francais du Cheval Anglo-Arabe (AA)	Anglo-Arab	3	6
Anglo European Studbook (AES)	Anglo-European	410	396
British Hanoverian Horse Society (BHHS)	Hanoverian	42	42
British Sport Horse Register (BSHR)	British Sport Horse	6	6
Belgish Warmbloedpaard (BWP)	Belgian Warmblood	643	687
Kladruber (CZECH)	Kladruber	4	2
Czech warmblood (CZEWB)	Czech Warmblood	12	11
Danish Warmblood (DWB)	Danish Warmblood	8	8
Verband hannoverscher Pferdezüchter (HANN)	Hanoverian	134	146
Verband Hessischer Pferdezüchter e.V.rmblood (HESS)	Hessian	8	5
Verband der Züchter des Holsteiner Pferdes (HOLST)	Holsteiner	611	604
Irish Sport Horse Register (ISH)	Irish Sport Horse	306	329
Koninkliijk Warmbloed Paarden Stamboek Nederland (KWPN)	Dutch Warmblood	1809	1897
Verband der Pferdezüchter Mecklenburg-Verpommern (MECKL)	Mecklenburger	13	13
The Dutch Riding Horse and Pony Studbook (NRPS)	Dutch horse & pony	65	63
Verband der Züchter des Oldenburger Pferdes (OLDBG)	Oldenburg	296	291
Rheinisches Pferdestammbuch e.V. (RHEIN)	Rheinlander	35	34
Royal Belgian Sports Horse Society (SBS)	Belgian Sports Horse	123	109
Selle Français (SF)	Selle Français	337	382
Sport Horse Breeding of Great Britain (SHBGB)	British Sport Horse	2	2
Studbook La Silla (SLS)	Mexican Sport horse	16	19
The Scottish Sports Horse (SSH)	Scottish Sports Horse	37	40
Swedish Warmblood Association (SWB)	Swedish Warmblood	12	10
Trakehner Breeders' Fraternity (TBF) Verband der Zuchter und Freund des Ostpreussischen Warmblutpferdes Trakehner Abstammung e.V. (TRAK)	Trakehner	59	57
Unione Nazionale Incremento Razze Equine (UNIRE)	Italian Horse	3	3
Westfälisches Pferdestammbuch e.V. (WESTF)	Westphalian	74	67
Studbook Zangersheide (ZANG)	Zangersheide	189	210
Other		50	44

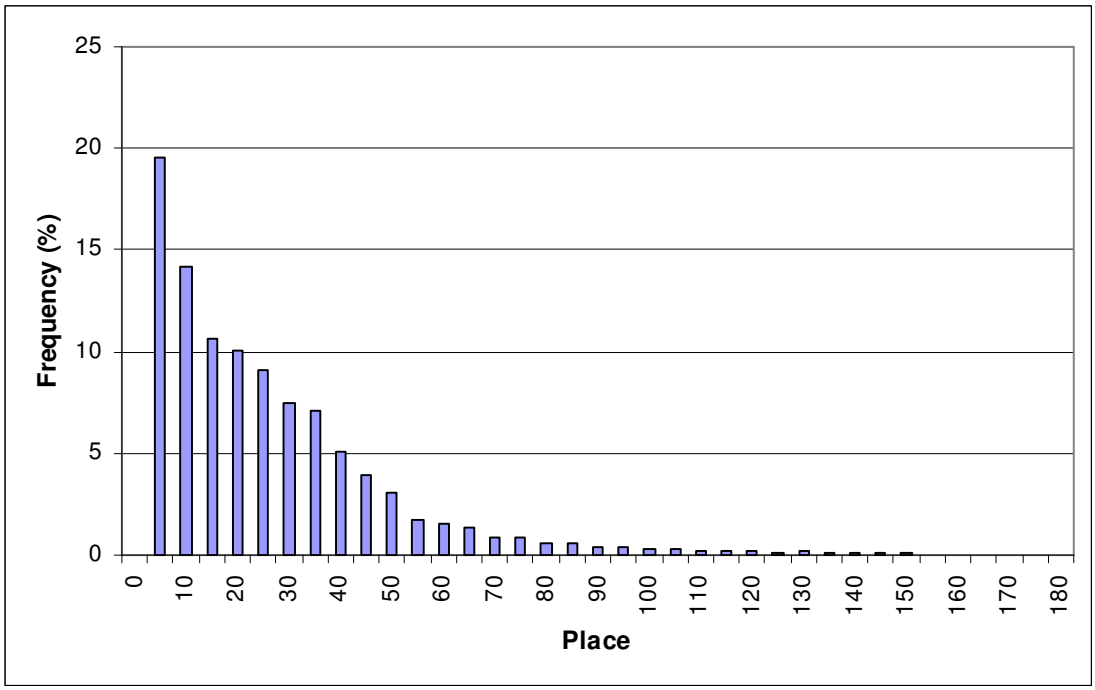


Figure 5.1a Distribution of place. Total no. records was 6239

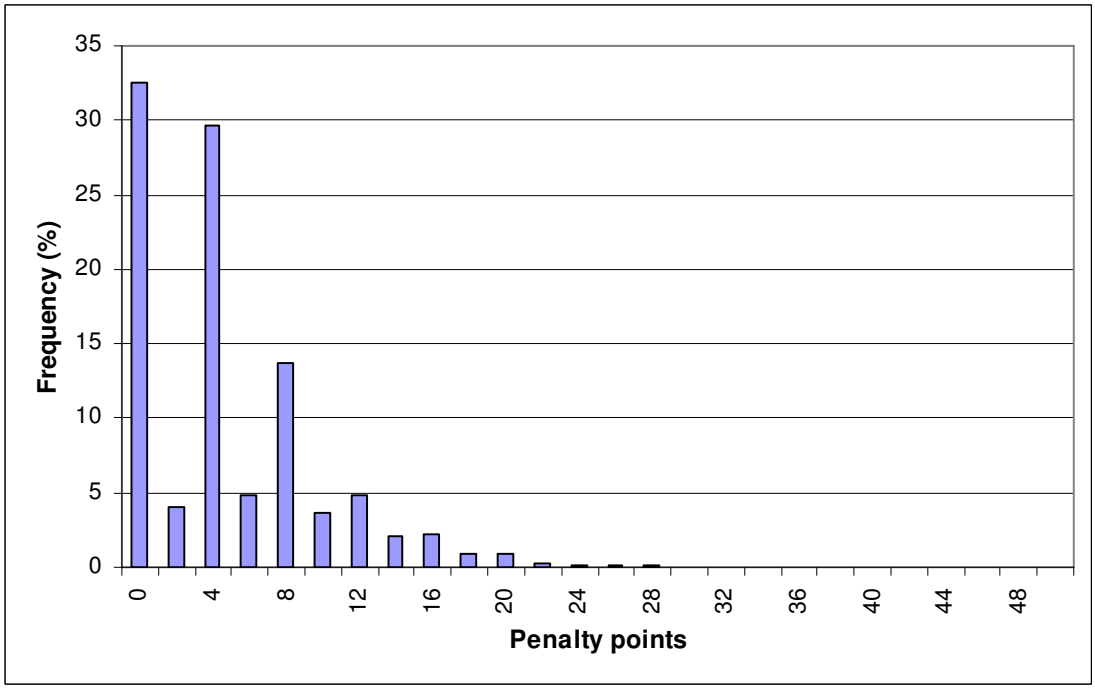


Figure 5.1b Distribution of penalty points. Total no. records was 6395

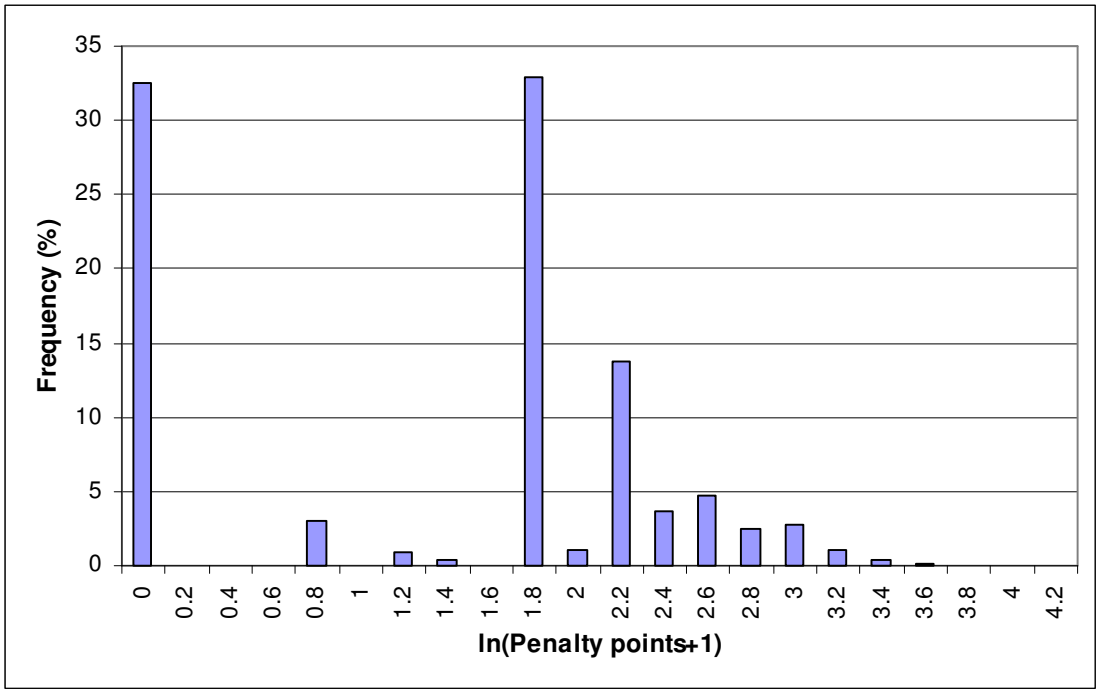


Figure 5.1c Distribution of natural log transformation of penalty points. Total no. records was 6395

5.4.2 Phenotypic parameters

The distribution of place, penalty points and the natural log of penalty points are shown in Figure 5.1, and a summary of the traits is given in Table 5.2. The untransformed distribution for penalty points ranged from 0 to 48 and was negative skew, with an accumulation of scores at 0. A natural log transformation was used, although the distribution remained relatively non-normal and there was an accumulation of scores at 0.

5.4.3 Genetic parameters

The variance component for the random effect of the horse, (the total variance due to the horse including both additive genetic and permanent environmental variance), as a proportion of the phenotypic variance, representing the repeatability which sets an upper limit to the heritability,

and the variance explained by classes within shows, again as a proportion of the phenotypic variance, are shown in Table 5.3. All were significant ($P<0.05$). The upper limit to the heritability was 8.9% in Model 1, and 5.4% in Model 2. The variance explained by the class within show was 2.6% and 6.5% respectively. The residual error variance was 88.5% and 88.1% respectively. This may include some variance due to the rider (which was excluded from the final model due to potential confounding). The total phenotypic variance by the two models was 0.97 for the normal score of place, and 0.90 for the natural log of penalty points.

Table 5.3. Variance components estimates as a proportion of total phenotypic variance, using place and penalty points as performance measures

	normal score of place			Ln(Penalty points)		
Variance component	**Component**	**s.e.**	**% of total phenotypic**	**Component**	**s.e.**	**% of total phenotypic**
Total horse (genetics + pe)	0.087***	0.012	8.948	0.049***	0.009	5.423
show - competing group interaction	0.025**	0.009	2.568	0.058**	0.011	6.453
Residual error	0.858***	0.018	88.483	0.791***	0.017	88.124
Total phenotypic	0.970		100.000	0.897		100.000

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

Table 5.4. Estimates of fixed effects

	normal score of place		Ln(Penalty points)	
	Estimate	**s.e.**	**Estimate**	**s.e.**
Height of jump	-0.0056**	0.0020	0.032***	0.002
Age				
2 – linear	0.0744	0.0599	-0.006	0.051
3 – quadratic	0.2259**	0.0816	-0.091	0.068

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

5.4.4 Fixed effects of age and jump height

Age significantly affected the mean of the trait, with a quadratic fit ($P<0.01$) for Model 1, but was not significant for Model 2 (Table 5.4.). The height of the jump also significantly affected the mean of the trait ($P<0.01$ and $P<0.001$ respectively). Note, as a good performance for penalty points will is represented by a low value of points, while a good performance for normal score of place is represented by a high place, the results for fixed effects have opposing signs in the two models.

5.4.5 Effect of studbook

For Model 3 (place), in general, there were no significant differences in performance from the most represented studbook – KWPN. However a better performance was observed by horses from the Scottish sports horse studbook. For Model 4 (penalty points), there were no significant differences in performance found between records from KWPN and other studbooks.

5.4.6 Effect of number of horses with clear rounds and number of horses with time faults

In a separate model with place as the performance measure (Model 5), the number of horses with clear rounds significantly improved the performance of the horse ($P < 0.001$), and the joint effects of the horse's genetics and permanent environment were found to account for 9.9% of the total phenotypic variance. In Model 6, however, the number of horses with time faults was not significant.

Using penalty points as a performance measure, in separate models (Models 7 and 8), both the number of horses with clear rounds (improving performance), and the number of horses with

time faults (detrimental to performance) were significant effects on the performance ($P < 0.001$). Respective repeatabilities were 5.6% and 5.7%.

5.4.7 Effect of the ground conditions and number of horses with clear rounds (2008 data only).

In a model fitting the percentage of horses that performed a clear round and the ground conditions, for 2008 data only (on 1294 records), the number of clear rounds was found to have a significant positive effect, while there was no significant effect of ground conditions. The repeatability was found to be 14.1%.

5.5 Discussion

This study attempted, for the first time, to perform genetic evaluations for British showjumping horses. However, a lack of available competition and pedigree data severely limited the study. An upper limit on the heritability of performance was estimated at 8.9% or 5.4% using performance measures of the place / penalty points respectively. These estimates are low compared with those from foreign studbooks, which range from 5% to 28% for the genetic effect alone (the higher estimate derived from aggregate measures of performance). However, results from the current study are an indicator of the presence of genetic variance, and in particular, the high degree of selection in the data may have produced a much reduced estimate of genetic variance. Evidence for higher estimates comes from the showjumping phase of BE (Kearsley *et al*., 2008; Stewart *et al*., 2010). The presence of significant genetic variance would give the potential for genetic progress to be made for showjumping performance. However, to perform accurate genetic evaluations, enabling selection for showjumping performance, it is necessary that more substantial data sources are made available.

Although the introduction of increased recording of variables for international and some classes of national data was a good initiative, unfortunately it has not been maintained. The volume of records has reduced over time, from what was already a minimum requirement. There was selection in the data recorded initially, in the fact that only international results and limited national competitions were recorded. In the latter years, selection increased, to encompass only data on horses that were consistently performing well. This was due to resource limitations in the BS. This limits the validity of the data both for genetic evaluations and for other analyses/ investigations, as suggested by the BS in 2007, to characterise horse's performance.

This large degree of selection will produce downwardly biased estimates of genetic variance and heritabilities. In addition, estimated breeding values can then only be estimated for a selection of horses. Internationally, genetic evaluations are based largely on national data, and many evaluations do not, or have not until recently, consider international performance data (Quinn, 2005). Many countries record only national performance data. The inclusion of international data would aid in estimating breeding values for competition at that level, which is a major aim of most breeding organisations (Koenen & Aldridge, 2002). There is often some selection in the recording of competition results in foreign evaluations, as frequently only horses that score points are recorded, or those that have earnings from winnings. It may be that the inclusion of international results, when the competition is held abroad, will have little effect on the estimated breeding values, as there may be relatively few national horses competing against each other, and therefore the information obtained on relative performances is scarce (Quinn, 2005). However, they may be easily included in evaluations, although most national evaluations do not include international results. National competition data represents the largest volume of records and the entire volume of competing horses, although this is already a sample of the population (depending on the methods of data recording and on the proportion of horses within the

population that compete). Therefore, the accuracy of an evaluation based on national data will be greater and estimated breeding values can be produced on a larger volume of the population, compared to evaluations based on international data.

Other data is collected by the BS. This includes the horse membership number, name, sire, dam, and performance measures including prize money, double clears, placing, date of birth and breed. Unfortunately, there were problems in getting access to this data, which hopefully will be resolved in the future.

Genetic evaluations have now been performed for dressage, eventing and young horse traits, recorded in GB (Stewart *et al*., 2010, Kearsley *et al*., 2008, Stewart *et al*., in submission a & b), and the initiative is that estimated breeding values, estimated for eventing, are to be published in the near future. International estimates, and those from BE, suggest that the genetic variance for showjumping is greater than the genetic variance for dressage, and therefore, quicker genetic progress can be obtained for this discipline, compared to dressage. Genetic correlations between showjumping and dressage are estimated to be low and either positive or negative (Thorén Hellsten *et al*., 2006; Ricard *et al*., 2000). Therefore, selection for dressage performance should have little effect on showjumping performance (although there is slight evidence that the genetic progress may be negatively correlated).

This analysis excluded horses that were eliminated, retired, etc. In addition, for penalty points, some classes were excluded, where the points were effectively awarded for good performance. However, it would be possible to include these results, at least for some performance measures. For the former, ranking could be used, and excluded values ranked lower than the lowest of

those that completed the competition. This was performed in evaluations for the Irish sport horse (Quinn, 2005).

Accurate pedigree, with as many ancestral generations as possible is required. In addition the use of a unique identifier for the name of the competing horse, and for the horses in the pedigree is also required. For this analysis, pedigree information comprising sires, dams and dams' sires was available, but only recorded by name. Missing or incorrect pedigree will reduce the accuracy of estimated breeding values, and thus the genetic gain (Woolliams, 2006), as EBVs will become regressed toward the mean, particularly for very high or low EBVs (Phocas & Laloë, 2004).

There were two performance measures available in this analysis – the place and the penalty points. The place can be converted to a normal score, and thus made suitable for analysis. The penalty points measure presents more difficulty however. Traits available for performance measures in horses are frequently problematic for analysis due to their non-normal distributions. Typical measures include earnings, penalty points etc... , which may be an aggregate measure, or one available for the individual competition. In this analysis, the penalty points measure produced a lower estimate of heritability. To some extent this may be due to the exclusion of records for classes where penalty points were not comparable. However, the non-normal distribution of penalty points reduces the validity of that model.

Genetic evaluations for showjumping are common in foreign populations. Other research has also examined genetic correlations between showjumping traits in a number of European countries (Denmark, France, Belgium, Ireland and Sweden) (Ruhlmann *et al*., 2009b). Genetic correlations were high, particularly between Belgium, France and Sweden. This indicates that similar traits are being selected for across countries. In Ireland, correlations between levels of

competition were examined; the estimate between low and high (national competitions) was 0.69, low and medium 0.97 and medium and high 0.83 (Aldridge *et al*., 2000).

5.6 Conclusion

The attempt to perform genetic evaluations for performance in show jumping competitions in GB was limited by a lack of competition and pedigree data. In the absence of being able to perform a genetic evaluation, and estimate the genetic variance component, an upper limit to the heritability of performance was estimated, at 8.9 or 5.4% depending on the performance measure. These estimates may be biased downwards due to the high degree of selection in the recording of the competition data. Significant genetic variance would indicate that selection for competition performance can be performed. In order for genetic evaluations to be performed, a greater volume of performance data, subject to less selection in recording of results is required.

6. General Discussion

6.1 Initial motivations of the work

Recent initiatives within the UK equestrian industry, notably by the BEF have been aimed at establishing more co-ordinated breeding programs for sport horses within GB. Genetic evaluations and the routine production of estimated breeding values are commonly performed by breeding organisations on the continent, and are performed by the studbooks that are most successful in producing horses which excel in elite competition. In light of this, the first academic study addressing the performance of genetic evaluations of sport horses in GB was commissioned in 2004 by the BEF (Kearsley, 2007). While Kearsley *et al.* (2008) successfully addressed the evaluations for eventing, the study was purely academic, and has not yet resulted in the publication of estimated breeding values. It was felt that further work to the models was required. This thesis continues the work on genetic evaluations for eventing and young horse evaluations but uses larger datasets for both, and develops further models, as well as addressing new work on competition disciplines for dressage and showjumping evaluations.

6.2 Resume of work

Genetic evaluations were performed for the competition disciplines – dressage, showjumping and eventing - and for traits recorded in young horses. Genetic evaluation systems for eventing are rare internationally, and evaluations for the other disciplines had not been performed in GB previously. Heritabilites were significant, but low for performance in dressage (0.15 / 0.10 without and with accounting for breed respectively) (Chapter 1) and for the dressage (0.072 – 0.090) and showjumping (0.089 – 0.162) phases of eventing (Chapter 2). An upper limit on the heritability for performance in showjumping was 0.089, although this was estimated using a

selection of very elite competition results, and thus was likely to be biased downwards (Chapter 4), which was corroborated by the higher estimate from the showjumping phase of the eventing evaluation (Chapter 2). Heritabilites for traits recorded in young horses were moderate and significant, ranging from 0.203 – 0.422 (Chapter 3). Thus, the study demonstrated that there is genetic variance for competition performance in the sport disciplines and importantly in the young horse traits and it should be possible to achieve genetic gain through selection.

The effect of breed on competition performance was explored using the dressage competition data, and was found to have a significant effect on performance (Chapter 1). Breed effects may be of particular importance in the GB population, compared to those on the continent, due to the relatively heterogeneous composition of our competing population. The data was not available to examine this directly in most other disciplines, although it is possible that similar effects will be seen.

The variance in competition performance attributable to the rider was examined for eventing (Chapter 2). Evaluations on the continent are unable to account for this important effect, due to the fact that data is generally not recorded. The proportion of phenotypic variance due to the rider was 0.113 - 0.291, which demonstrates the relative importance of the rider in competition, particularly for the dressage phase and the potential benefit from properly accounting for this source of variance in evaluations.

Novel methods to estimate large (co)variance matrices were investigated (Chapter 3). Two methods that combined sub-(co)variance matrices from bivariate matrices were found to be most appropriate, and suitable for application. A third method, newly implemented in ASReml, that performed a large multivariate analysis by splitting the model into a number of submodels, and

iterating between them, was not entirely reliable and requires more work before use. It is a potentially a valuable tool should its unreliability be overcome.

Genetic correlations between and within competition disciplines were estimated for the phases of eventing competition (Chapter 3). Correlations between grades within dressage and within showjumping were significant and high, indicating that selection for performance at lower levels of competition also selects for performance at more advanced grades. Correlations between disciplines were generally not significantly different from 0, and if this were so, then selection for change in one phase does not necessarily promote changes in other phases. Data other than eventing data were not available to estimate genetic correlations between the competition disciplines themselves, however the eventing work may act as some evidence for the GB population, where correlations between disciplines were generally not significant, and where significant were low and at low grades of competition. The GB Futurity results contradict this, finding very high genetic correlations between disciplines; however, this is thought to be due to the very young age at evaluation, at which point differentiation between disciplines has not yet occurred.

The young horse tests have been successfully established in GB. Heritabilites of traits recorded in young horses are higher than for competition performance, reflecting reduced environmental variance in young horse tests compared to adult competition, and the exclusion of a rider effect. The high heritabilites indicate that if competition performance is genetically correlated to later competition performance, then the young horse data will be very valuable for genetic evaluations. Integrated evaluations will likely be the most appropriate measure in the future. The use of young horse data will decrease the generation interval, providing EBVs earlier in life and greater opportunity for genetic gain.

The study brought to light some interesting aspects relating to current breeding practice in GB. There was evidence of a genetic gain of 0.046 genetic standard deviations per year between 1985 and 2001 for performance in dressage under the current breeding programs. This is encouraging evidence, and most likely reflects the fact that the introgression and increased mixing of genes from the continent is occurring in GB (Chapter 1). For eventing, the genetic variance uncovered for cross-country phase was very low (or was unable to be elucidated given the data) (Chapter 2). By contrast, the heritabilites for the other phases were much higher. Breeding practice in GB tends to use the cross-country abilities of a stallion for selection for competition performance in eventing. Evidence from this study suggests that greater genetic gain and improvement in competition performance will be achieved by selecting for one of the other phases of competition, and highlights the relevance of the study.

6.3 General limitations

This study was successful in performing genetic evaluations for academic use in GB, and thus made headway into their implementation and the routine production of genetic evaluations. There are, however, caveats to the analysis, for instance relating to the data that has been collected by the equestrian industry in the past and present, and, therefore is available for analysis.

6.3.1 Data Availability

There are important problems in the data for the purposes of genetic evaluations. The problems are primarily concerned with the data providers. It must be remembered that in most instances, the data has been collected for alternative purposes by the competition bodies (e.g. recording horses that scored points). These bodies have resource limitations and in many cases rely on voluntary contributions.

6.3.1.1 Competition data

Substantial datasets of competition results were available from the competition bodies for dressage and eventing, dating back to 1978 and 1999 respectively. However, as is commonly a problem internationally, there was selection in the recording of the dressage results (only records where scores were > 60% were recorded as a general rule) which at present downwardly biases heritability estimates. Recording all records, irrespective of score, would add to the value of the dressage dataset for genetic evaluations. For showjumping, the data available was very highly selected (as it included only a selection of international results), and a less selected dataset, including national data, needs to be made available in order to perform adequate evaluations. In addition to selection in recording of data, there will also be selection in the horses that compete. Selection among young horses that are evaluated is likely to be greatest. The young horse dataset was limited in size due to the recent establishment of the scheme. Encouragingly participation in the scheme has been very successful, and uptake increases annually, indicating that data will accumulate quickly over time to form a very valuable dataset. In the future, the most successful analysis will likely be performed using competition data together with young horse data, as is now performed by many continental studbooks.

In the long-term, if the data is to be used for genetic evaluations, certain aspects must be rectified. A major issue is the individual identification of competing, or evaluated horses. For some datasets (namely eventing and dressage) competing horses were identified using competition ids, which reliably uniquely identify the horse within that dataset. Other, unique identifiers such as the id assigned to the horse by the competition body, the passport number, or other identifiers may have been recorded, but very inconsistently to the extent that they cease to be of any use. In others (showjumping and the Futurity) horses are recorded only by name. Given the fact that there are potentially spelling errors in names, names are shortened in different

ways, and that many horse names are similar (particularly for mares), this makes errors likely. In identifying horses for the current analysis, a lot of "considered research interpretation" was required, which introduces the potential for errors.

A unique identifier that is consistent across competition bodies is required. Both the UELN and the passport number, which have been introduced recently (2000 and 2004 respectively) present these, although to be of use one at least must be used consistently. Recent changes in the systems of data recording by the competition disciplines have introduced the recording of passport numbers with the competing horse's name for competition disciplines (Jan Rogers personal communication, March 2011). However, the newly introduced recording of young horse evaluations, and international showjumping datasets lack unique ids. Given the importance of the accurate recording of horses, it is essential that these are established as soon as possible to add to the infrastructure of future data.

The competition bodies have recently established a new joint recording scheme, that records the passport number as a unique identifier with competing horses. From this point onwards, data should be consistently recorded across disciplines. This new recording scheme should also facilitate the recording of horses so that they may be uniquely and consistently identified. It also presents an opportune time to further consider the uses of the data, and potential improvements in recording, for instance, i) genetic information that can be included – at present this may be limited to pedigree, however, genotyping is a potential future development ii) fixed effects that may be considered in a genetic evaluation iii) whether it can be linked to other databases – equine (e.g. NED) or other and iv) commercial activities.

The lack of a unique identifier (for competing horses and pedigree) also has implications for, and hinders, the international comparisons of evaluations. This was experienced in a study by Thorén Hellsten *et al.* (2008), while assessing connectedness between some of the major European studbooks. Stallions in different countries were often found to have different identifiers and names differed due to inclusion/ exclusions of prefixes and suffixes.

6.3.1.2 Pedigree data

The pedigree data available for the study was limited with respect to the number of generations that were available and accuracy of recording. Pedigree data available for the study was, at most, the sire, dam and dam's sire of competing horses. Recording was generally only using the name. In some cases, other identifiers, such as passport numbers had been used, but with insufficient consistency to be of any use. There were also errors in recording of horse names. In many instances, only the sire data (names) were distinguishable enough to be used. In the present work, although dam was recorded, it was often not sufficiently well recorded for use. Only names were available, and dam names tend to be similar or pet names, thus it was impossible to distinguish dams by name alone. The use of names required manual matching with "considered research interpretation", i.e. the considered matching of names given the data, the potential for errors in recording, and information available from external sources such as internet searches. This was all based on the side of caution due to the fact that errors in pedigree have potentially twice the detrimental effect on genetic gain as missing pedigree (Woolliams, 2006). It was also a very labour intensive and time consuming part of the study. Alternative methods such as computational programs to match based on similarity of names were considered. However, due to the presence of pre-fixes, suffixes, and to the similarity of many names for different horses this was deemed inappropriate.

For the purposes of reliable genetic evaluations - and for numerous other potential uses within the equine industry - unique identifiers are required for the recording of pedigree. The consistent use of one, or both of the UELN and passport numbers would provide this. It is important that recording of these is established as soon as possible, in order to establish an infrastructure. It may never be possible to introduce unique identifiers in retrospect, as this risks the introduction of errors into the data. In addition, this makes identification of stallions between foreign countries difficult, with different names or identifiers used by different foreign studbooks (Thorén Hellsten *et al*., 2008). Very recent political upheavals in the UK over passport issuing authorities may further exacerbate the problems in allocating unique identification. The original 60 passport issuing authorities may now be reduced to one, which in itself would be an advantage, but the transition period may well produce additional problems. The recent changes to data recording systems of the competition bodies, as mentioned above, now record the passport number with the competing horse's name, however, this is not necessarily true for the pedigree of the horses.

Historically, competition bodies have recorded results data for purposes not specifically related to the pedigree. Given resource limitations, the recording of pedigree has not been of primary importance. One aim of the National Equine Database, integral to its other objectives, was to assemble pedigree data from various sources such as studbooks. However, data available to this study from this source was limited, possibly due to problems with access to data originating from the studbooks. In common with this study, the National Equine Database is likely also to have experienced problems in assembling pedigree data due to identification of horses.

Due to these limitations, in many instances only sire models could be used, as opposed to the preferred animal model. This will have affected the analysis in that the animal model is regarded

as being more accurate for the estimation of variance components and estimated breeding values. More genetic connections are accounted for and the animal model can overcome problems due to selection within the data. The sire model assumes that sires are randomly mated, does not account for relationships between sires (or between dams or between sire and dam) and assumes no inbreeding or selection as well as constant genetic variance over time. In horse populations, many of these assumptions do not hold. Selection has been practiced over time and non-random mating occurs, in that superior horses tend to be mated as do inferior horses. Thus with a sire model, EBV estimates for superior sires (mating with superior mares) will be overestimated, and for inferior sires (mating with inferior mares) will be underestimated. The extent to which EBVs will be in error will be greater for sires that are further from the mean, i.e. the very best and worst of sires, and greater for traits with greater heritabilities. In practice the model used is usually determined by the pedigree available. Within the current study, any dam half sib groups would be small, with few progeny compared to the sire. However, including dam pedigree in the analyses could have contributed some genetic connections to the analysis, which would have increased its accuracy.

The lack of depth of pedigree will mean that non-random mating in the population is not accounted for, which it is by large pedigrees. In many foreign evaluations, the depth of pedigree available is much greater. Pedigree is sourced from the studbooks, which have a clear interest in the pedigree, and indeed correct pedigree recording may be necessary for eligibility for entry into the studbook. For instance the SWB includes at least 5 or more ancestral generations (Thorén Hellsten *et al*., 2009b) and the Hanoverian 8.43 generations (Hamann & Distl, 2008). The completeness of pedigree also has a large effect on genetic evaluations and EBV estimates. In a study by Thorén Hellsten *et al.* (2009b), in a 5-generation pedigree, sires/ stallions with limited genetic connections to the dataset (foreign sires) had changes in breeding value estimates

of approximately 1 genetic standard deviation. Inaccuracies in data recording will cause less accurate estimates, and less genetic gain.

6.3.2 Limitations in the distributions of traits

Traits recorded as performance measures for sport horses are typically not ideal for genetic evaluations. Frequently traits are recorded as aggregate measures, such as annual or life-time measures, and traits include penalty points, earnings, place or ranking. In this study, the traits available for analysis were: for eventing the penalty points, for showjumping the place and penalty points, for dressage the percentage (less than 60% censored) and for Futurity 0 – 10 score. In the case of the use of penalty points in the showjumping data, these were not normally distributed, and there was a large aggregation of scores at zero. Transformations such as square root, and log were inappropriate, and did not achieve a more appropriate distribution. Other transformations could be investigated. For instance, penalty points could potentially be transformed into a categorical trait, and analysed using a (multi)threshold generalized linear mixed model. These models are characterized by the use of a link function which allows the analysis to be performed on the underlying scale. Typical link functions are the probit and the logit. Results of such an analysis would allow the comparison of thresholds across categories of sires or shows or grades. These 'making-do' approaches are not ideal, substituting categorical traits for continuous traits, with the resulting loss of information.

6.4 Future developments in genetic evaluations

Developments in genetic evaluations and in related aspects of genetics are important to consider, both for pure interest, and with a view to planning the infrastructure that may be required. Most studbooks breed not just for competition performance, but also for measures of aesthetics, conformation, health, character/ behaviour and fertility. The importance of these traits is

reflected in their breeding goals, for instance, for the SWB the breeding goal is "A noble, correct and durable sport horse which through its temperament, rideability, good movements and/or jumping ability is expected to be internationally competitive in dressage, showjumping or eventing" (Janssens, 2008). Young horse tests present an ideal opportunity at which to introduce less traditional measures and assess their value. Now many foreign evaluations use measures of health and character and behaviour scored at young horse tests (Thorén Hellsten *et al*., 2006). For example, in the SWB, health traits health (medical) and health (orthopedic) are recorded in young horse tests, and in the Oldenburg, and Hanoverian, character, temperament, willingness to work and constitution are recorded (Janssens, 2008). The GB Futurity young horse tests also include a measure of temperament and of veterinary traits, reflecting health. Longevity is a further trait that may be of interest for inclusion in evaluations, and potentially may be quantifiable, as phenotypic associations between young horse tests and longevity suggest that young horse tests may be useful in prediction of longevity (Wallin *et al*., 2001).

However, inclusion of new traits in genetic evaluations, such as health, character and longevity is encumbered by complexities of definition, measurement and lack of available data. For instance, the temperament of a horse for performance within the discipline has a great effect on performance, and potentially is one of the most important characteristics. However, the measurement of temperament is complex. Various tests have been developed for research purposes (Houpt & Kusunose, 2000), and there are generally accepted breed differences in temperament, thus demonstrating a genetic component (Houpt & Kusunose, 2000; Hermsen, 1997 as cited by Houpt & Kusunose, 2000). Temperament assessments are made at young horse tests, however, these may be subjective and limited in time. This assessment represents a snapshot potentially reducing the repeatability of the assessment. Factors that may affect the horse on the day, i.e. due to temporary environment for instance the distance travelled to the

evaluation centre, whether the horse has attended an evaluation before and, for young foals, past training and handling may introduce biases.

Other opportunities for the collection of data may be possible, for instance, from veterinary clinical work, insurance data (Wallin *et al*., 2000), auction sales and riding schools. Studies have used insurance data to assess the causes of death of animals (Clausen *et al*., 1990; AGRIA, 1995; Heisele, 1995, all as cited by Wallin *et al*., 2000) and radiographic (X-ray) data from auctions was used by Stock *et al*. (2008 and 2005) to assess joint/ bone pathologies such as osseous fragments and deforming arthropathy, which were then considered for inclusion in genetic evaluations for research purposes (Stock *et al*., 2008). Evidence suggests that genetic correlations with competition performance are slight or zero however, and thus although simultaneous selection could be performed, indirect selection is not viable (Stock *et al*., 2007). GB passport issuing authorities also have the potential to collect data. Some research studies have used questionnaires (Wallin, 2000). However, as with the use of any data not collected for a specific purpose, there may be biases or selection in the data, which may be available for only a limited proportion of the population. The UK passport system records the death of horses, and thus in the future will present the opportunity to capture longevity data.

There are many serious health problems in horses, and evidence of a heritable component to these would make their inclusion in genetic evaluations of value. Diseases known to be due to genetic effects include polysaccharide storage myopathy, (Valberg, 1997 as cited by Collinder & Rasmuson, 2000; Horse Genome Project, 2007) and osteochondrosis (Dierks *et al*., 2010). Other traits present may present financial implications for breeding, for instance colour. Data could be collected on these traits, with a view to inclusion in genetic evaluations in the future.

6.5 The Equine Genome Sequence

Whilst there are many possibilities and aspirations, and there is a need to generate momentum to overcome the infrastructure limitations and prompt action. The sequencing of the horse genome in 2007 (Wade *et al*., 2009) presents new ways to generate momentum through the opportunity to examine associations between traits such as performance, health and character with regions of the genome, providing the phenotypes are recorded. These include athletic performance, behaviour and fertility and reproduction (Horse Genome Project, 2007). Genome wide association studies have begun to examine quantitative trait loci (QTL) for instance for osteochondrosis (Dierks *et al*., 2010, Dierks *et al*., 2007; Horse Genome Project, 2007), and for racing performance in the thoroughbred, which identified a candidate region close to the myostatin gene (Tozaki *et al*., 2010). Once identified, information gained from the outcomes of the sequencing of the genome could be combined to be used in marker assisted selection (MAS), which has been applied in other livestock species, for instance in cattle.

Other novel technologies in animal breeding may also have potential applications in horse populations. Genomic selection uses information on QTL effects on a trait, assessed in a large sample of the population with genotypic and phenotypic measurements, to then predict breeding values, for a smaller sample of the population, using only genotypic data (Meuwissen *et al*., 2001). Large scale genotyping is required. It is of interest to consider the infrastructures/ systems for arranging large scale genotyping or single nucleotide polymorphism (SNP) estimates in a sample of the GB, or continental, population. The most appropriate time for collection of information from a reference population would likely be young horse tests, where genotyping and phenotypic information could be collected simultaneously. The size of the reference population however – for instance an estimated 6000 in the Dutch population (Ducro, 2011) –

will require some time to establish, and is an inhibition to introduction of the technology, even abroad, within the next few years. Compared to other livestock, there are distinctions in the life and breeding programs of horses which mean that genomic technologies could potentially have a great impact. Long generation intervals (typically 8 – 12 years), the gelding of a large proportion of the male population before selection traits become available, the fact that some of the primary selection traits, i.e. adult competition performance, only become available late in life and difficulty in measuring traits, currently greatly limit the genetic progress that can be made. Genomic selection could overcome many of these problems, reducing the generation interval and increasing genetic gain (whilst maintaining genetic variation and decreasing inbreeding effects).

6.6 What can be learnt from young horse tests abroad?

Genetic evaluations of sport horses are long running on the continent. The various types of young horse tests are long established, and knowledge from these can be used to inform the best procedures for the Futurity scheme. The precedent for this was set in the fact that the current Futurity scheme was initially modelled on the two young horse tests in Sweden.

Foreign tests have traditionally been of fairly long duration, for example stallion tests at 3 to 4 years are commonly at least 60 days (Thorén Hellsten *et al*., 2006). In contrast, the Futurity is a very short test, conducted on one day (with a duration of minutes). However, evidence indicates that relatively short tests are genetically as good as tests of longer duration (Thorén Hellsten *et al*., 2006). Heritabilities are generally of similar magnitudes regardless of the duration of the tests (within groups of stallions, mares or young horses). German studies suggest that long (30 day) station tests for stallions are genetically equivalent to young horse field tests (Brockmann, 1998 as cited by Thorén Hellsten *et al*., 2006). Overall, the evidence suggests that short tests

(potentially with repetitions) are suitable (Thorén Hellsten *et al*., 2006). The evidence above gives some assurance to the Futurity scheme, although its duration remains one of the shortest tests. The value of so short a test still has to be established, and Futurity may benefit from repeated measures on the horses over an interval. Although more complex tests may be more powerful in differentiating between horse abilities, tests need not necessarily be too complex (Gelinder *et al*., 2002, as cited by Thorén Hellsten *et al*., 2006). The easier and shorter the tests are, the greater the participation may be, enabling greater selection intensity.

The age at which the Futurity assesses horses is very young (foals – 3 year olds) compared to young horse tests abroad (3 years and older). The earlier that the test occurs (giving a reasonable heritability) enables a shorter generation interval, and thus greater opportunity for genetic gain. On the other hand, to test at older ages enables assessment between disciplines with greater differentiation and more demanding tests. The age of Futurity assessment was set due to demand for and uptake of tests at this age, at which point there is little differentiation between the traits required for the different disciplines. In the Swedish tests, the high genetic correlations suggest that there is little difference between 3 and 4 years in this population (Thorén Hellsten *et al*., 2006). What remains to be established is the genetic correlation of 'Futurity' age with adult performance, since 3 years is an upper age of Futurity but a lower age of other schemes.

In order to enable genetic correlations to be assessed, some, if not all horses must be evaluated for multiple (both) disciplines at young horse tests. This is standard practice in most foreign evaluations. It is less feasible to assess using competition data, due to the fact that many horses will only compete in one discipline. In the Futurity scheme, horses can only be entered for one discipline. However, due to the young age at testing in the Futurity, evidence suggests that the genetic correlations between disciplines are not different from one, and thus there is little need

given the current system. However, correlations were estimated based on a dataset of limited size, and assumptions. If the tendency is not to assess horses for multiple disciples at one event, the data enabling us to re-asses this will take time to accumulate.

A further potential development in the Futurity is the exclusion of approved stallions, which is common practice in other countries. However, this will be detrimental to the accuracy of evaluations, introducing selection and biases to the evaluated population.

6.7 Wider significance of the thesis

This thesis demonstrated the utility of a novel tool to perform genetic evaluations on a large number of traits, and compute large covariance matrices with relatively small computing resources (Chapter 3.). This is a valuable tool in allied fields, such as animal breeding and evolutionary biology, where it provides an important advance in the current methods available.

In animal breeding, breeding programs are optimally designed using knowledge of the genetic-covariance matrix between traits that may be selected for directly or indirectly. Breeding programs and selection objectives may then be designed, for example based on optimal contribution or economic gain. Established routine genetic evaluations may already include large numbers of traits, for instance dairy cattle genetic evaluations can include 22 traits (Meyer *et al.*, 2004), and there is a demand for evaluations with up to 100 traits in forestry research in Sweden (B. Andersson personal communication December 2010). Yield traits were traditionally included in breeding programs, and others, such as conformation, disease resistance, longevity, fertility and reproduction have been incorporated over time. Traits other than production have become increasingly important (Powell & Norman, 2006), and evaluations may benefit from inclusion of these. For instance, dairy cattle fertility is in decline, and further measures of fertility, ideally

which can be measured in juvenile males and that are correlated with female fertility, such as hormone levels, would be ideal, as they overcome time delays and problems in measurement in females (Hayhurst *et al*., 2009). The inclusion of additional traits in genetic evaluations has been encumbered by computational requirements. Furthermore, many established genetic evaluation systems are based on outdated variance component estimates, due to the computation resources that have traditionally been required for re-estimation. A tool that can facilitate the computation of genetic evaluations with large numbers of traits will have important applications in animal breeding.

In evolutionary biology, our understanding of fitness and selection is informed by heritabilites and the genetic covariance between traits. The study of evolution is best performed in a multivariate context, which can provide greater insights into selection and genetic architecture than study with univariate analyses (Lande & Arnold, 1983; Philips & Arnold, 1989). Questions such as why there is there is genetic variance in traits in nature, yet directional selection under field conditions is often unsuccessful, may be best examined in this manner, for instance, by using the eigenvalues of matrices - as opposed to individual heritabilites and correlations - to provide insight about genetic constraints (Walsh & Blows, 2009; Blows 2007). The role of selection versus drift may also be examined from the evolution of the genetic (co)variance matrix (Arnold *et al*., 2008; Roff, 2000, Hine *et al*., 2000). The availability of a tool to produce large genetic covariance matrices will facilitate study into these areas.

6.8 Following work on genetic evaluations in GB

Due to time constraints, a full genetic analysis to assess genetic correlations between disciplines could not be performed. However, genetic correlations were estimated by correlating estimated breeding values. This exercise was performed for the analyses where BVs were estimated with

sufficient reliability (eventing and dressage). Correlations were estimated between dressage and each of the three disciplines of eventing (where EBVs across grades were averaged), and with a reliability lower threshold limit set at 0.1. The correlations between the dressage phase of eventing and dressage competition was 0.203 (based on 215 sires), between the showjumping phase of eventing and dressage was 0.079 (based on 381 sires) and between the cross-country phase of eventing and dressage was 0.232 (based on 15 sires). These figures suggest that genetic correlations between dressage competition and each of the phases of eventing are relatively low. However, for dressage and cross-country this was based on a small number of sires with progeny in both types of competition. Note also that many EBVs in the calculations had limited reliability.

For genetic evaluations in sport horses to be successful in GB the general public needs to understand and use the breeding indexes. As part of this thesis work, the intention is that articles publicising the eventing EBV work will be published in a number of journals including British Breeder and Horse and Hound. These will promote the publication of estimated breeding values and aid their acceptance and use in the public domain.

In addition to this thesis, a report to BEF recommending ways in which the routine implementation of estimated breeding values for eventing can be performed by the BEF has been prepared (Appendix 1). This report describes the process of taking competition data and using it to predict breeding values. It is intended for use by the BEF decision makers, and by individuals within BEF who will use the document for practical use.

6.9 Conclusions

The models have been developed but the infrastructure for evaluation and publicising the EBV (genomic or otherwise) needs to be developed further. Given the rapid advances of genetics and genomics breeding organisations in the UK need to resolve the future directions for breeding sport horses and put in place the necessary tools.

7. Bibliography

AGRIA, (1995). 90-94-Femårsskrift, AGRIA, Stockholm (Insurance Statistics 1990–1994).

Aldridge, L.I., Kelleher, D.L., Reilly, M. & Brophy, P.O. (2000). Estimation of the genetic correlation between performances at different levels of show jumping competitions in Ireland. *Journal of Animal Breeding and Genetics*, 117, 65–72.

Árnason, T. (1987). Contribution of various factors to genetic evaluations of stallions. *Livestock Production Science,* 16, 407–419.

Árnason, T. (1997). The selection intensity in Standardbred trotters in Sweden as measured by BLUP animal model index. *48th Annual meeting of the European Association of Animal Production*. Vienna, Austria.

Árnason, T. & Van Vleck, L. (2000). Chapter 17 – Genetic improvement of the horse. In: Bowling AT, Ruvinsky A (eds.) *The genetics of the horse*. CABI Publishing, Wallingford, UK, 411–438.

Arnold, S.J., Bürger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. (2008). Understanding the evolution and stability of the G-matrix. *Evolution,* 62, 2451-2461.

Blows, M.W. (2007) A tale of two matrices: multivariate approaches in evolutionary biology. *Journal of Evolutionary Biology,* 20, 8-15.

Bösch, M., Reinecke, S., Röhe, R. & Kalm, E. (2000) Genetische analyse von Meikmalen in der reitpferdezucht. *Züchtungskunde* 72, 161-171.

Brockmann, A. (1998). Entwicklung einer Eigenleistungsprüfung im Feld für Hengste unter Berücksichtigung der Turniersportprüfung. Dissertation. Universität Göttingen, Germany.

Celly (2004). The Warmbloods [online]. Available from: http://www.suite101.com/article.cfm/horse_talk/106955/2. [Accessed May 2011]

Clausen, M., Preisinger, R. & Kalm, E. (1990). Analyse von Krankenheitsdaten in der deutschen Warmblutzucht. *Züchtungskunde,* 62, 167–178.

Clayton, D. & Rasbash, J. (1999). Estimation in large crossed random-effect models by data augmentation. *Journal of the Royal Statistical Society. Series A,* 162, 425–436.

Collinder, E. & Rasmuson, M. (2000). Genetic Aspects of Disease in Horses. In: Bowling AT, Ruvinsky A (eds.) *The genetics of the horse*. CABI Publishing, Wallingford, UK, 157–170.

Cunningham, E.P., Dooley, J.J., Splan, R.K. & Bradley, D.G. (2001). Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Animal Genetics,* 32, 360–364.

Dierks, C., Komm, K., Lampe, V. & Distl, O. (2010). Fine mapping of a quantitative trait locus for osteochondrosis on horse chromosome 2. *Animal Genetics,* 41, 87–90.

Dierks, C., Lohring, K., Lampe, V., Wittwer, C. & Drogemuller, C. (2007). Genome-wide search for markers associated with osteochondrosis in Hanoverian warmblood horses. *Mamalian Genome,* 18, 739–747.

Dietl, G., Hoffmann, S. & Reinsch, N. (2005). Impact of trainer and judges in the mare performance test of Warmblood Horses. *Archiv Tierzucht*, 48, 113-120.

Dubois, C., Manfredi, E. & Ricard, A. (2008). Optimization of breeding schemes for sport horses. *Livestock Science,* 118, 99–112.

Dubois, C. & Ricard, A. (2007). Efficiency of past selection of the French Sport Horse: Selle Français breed and suggestions for the future. *Livestock Science,* 112, 161–171.

Ducro, B.J., Koenen, E.P.C., v Tartwijk, J.M.F.M. & v Arendonk, J.A.M. (2007). Genetic relations of first stallion inspection traits with dressage and show-jumping performance in competition of Dutch warmblood horses. *Livestock Science,* 107, 81–85.

Ducro, B.J., Bovenhuis, H. & Back, W. (2009). Heritability of foot conformation and its relationship to sports performance in a Dutch Warmblood horse population. *Equine Veterinary Journal,* 41, 139-143.

Ducro, B.J. (2011). Relevance of test information in horse breeding. PhD Thesis. Wageningen University, The Netherlands.

Falconer, D.S. & Mackay, T.F.C. (1996). *Introduction to quantitative genetics*, 4th edn. New York, NY: Longman Group Scientific & Technical.

FEI (2008). Short list for dressage [online]. Available from: http://www.fei.org/olympics/References/Qualifications/Documents/Dressage%20Final%20Entries-31July08.pdf. [Accessed January 2009]

Fredricson, I. (2003a). The Swedish Sport Horse [online]. Available from: http://www.albertahorseindustry.com/hboc/2003/proceedings/swedish_sport.pdf. [Accessed July 2011]

Fredricson, I. (2003b). An international sport horse database in progress [online]. Available from: http://www.albertahorseindustry.com/hboc/2003/proceedings/database.pdf. [Accessed July 2011]

Gelinder, Å., Thorén, E., Näsholm, A. & Philipsson, J. (2002). Relationships between test results at 3 and 4 years of age for Swedish riding horses. *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, Montpellier, France, August. pp. 0-4.

Gerber, E., Sjöberg, A., Näsholm, A. & Philipsson, J. (1997). Genetic parameters for conformation traits of warmblood horses in Sweden. *48th Annual meeting of the European Association of Animal Production*. Vienna, Austria, 25–28 August. pp. 378.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2006). ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009a). ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, UK www.vsni.co.uk.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009b). ASReml 3.0 html help – Imputation [online]. Available from: http://www.vsni.co.uk/software/asreml/htmlhelp/asreml/impute.htm. [Accessed May 2011]

Hadfield, J.D., Wilson, A.J., Garant, D., Sheldon, B.C., & Kruuk, L.E.B. (2010). The misuse of BLUP in ecology and evolution. *The American Naturalist*, 175, 116 – 125.

Hamann, H. & Distl, A.O. (2008). Genetic variability in Hanoverian warmblood horses using pedigree analysis. *Journal of Animal Science,* 86, 1503–1513.

Hayhurst, C., Flint, A.P.F., Lovendahl, P., Woolliams, J.A. & Royal, M.D. (2009). Genetic variation of metabolite and hormone concentration in UK Holstein-Friesian calves and the genetic relationship with economically important traits. *Journal of Dairy Science*, 92, 4001-4007.

Heisele, C. (1995) Schadenfälle in der Langfristigen Pferdelebensversicherung unter berucksichtigung der tierärtlichen Mitwir kung in der Pferdelebensversicherung. Thesis, University of München. München, Germany.

Henderson, C.R. (1973). Sire evaluation and genetic trends. Proceedings of the animal breeding and genetics symposium in honor of Dr J L Lush. *American Society of Animal Science*, Champaign, IL pp. 10-41.

Higham, N.J. (2002). Computing the nearest correlation matrix – a problem from finance. IMA *Journal of Numerical Analysis,* 22, 329–343.

Hill, W.G. & Thompson, R. (1978). Probabilities of non-positive definite between-group or genetic covariance matrices. *Biometrics,* 34, 429–439.

Hine, E., Chenoweth, S.F., Rundel, H.D. & Blows, M.W. (2009). Characterizing the evolution of genetic variance using genetic covariance tensors. *Philosophical Transactions of the Royal Society B,* 364, 1567 – 1578.

Horse Genome Project (2007). Some Definitions of Genetic traits and conditions currently under study [online]. Available from: http://www.uky.edu/Ag/Horsemap/hgpprojectdefs.html. [Accessed July 2011]

Huizinga, H.A., Boukamp, M. & Smolders, G. (1990). Estimated Parameters of Field Performance Testing of Mares from the Dutch Warmblood Riding Horse Population. *Livestock Production Science,* 26, 291–299.

Huizinga, H.A. & Vandermeij, G.J.W. (1989). Estimated parameters of performance in jumping and dressage competition of the Dutch Warmblood horse. *Livestock Production Science,* 21, 333–345.

Huizinga, H.A., van der Werf, J.H.J., Korver, S. & van der Meij, G.J.W. (1991) Stationary performance testing of stallions from the Dutch Warmblood riding horse population: 1. Estimated genetic parameters of scored traits and the genetic relation with dressage and jumping competition from offspring of breeding stallions. *Livestock Production Science,* 27, 231–244.

Houpt, A.H. & Kusunose, R. (2000). Genetics of Behaviour. In: Bowling AT, Ruvinsky A (eds.). *The genetics of the horse*. CABI Publishing, Wallingford, UK, 181 – 306.

Hermsen, J. (1997). *The Horse Encyclopedia*. Firefly Books, Buffalo, New York.

Interstallion (2005). Recommendations on choice of scale and reference population for publication of breeding values in sport horse breeding. Available from: http://www.biw.kuleuven.be/genlog/livgen/docs/publicationscale.pdf. [Accessed May 2011]

Janssens, S. (2008). Breeding programs and estimated breeding values [online]. Katholieke Universiteit Leuven. Available from: http://www.biw.kuleuven.be/genlog/livgen/chgs_is_Breedprog.html# [Accessed September 2008]

Jorjani, H. (2000). Well-connected, informative sub-sets of data. *Proceedings of the 2000 Interbull Meeting. Bled, Slovenia*, May 14–15. *Interbull Bullentin*, 25, 22–25.

Kaiser, M., Duda, J. & von Butler-Wemken, I. (1991). Genetische und nict genetische Einflüsse auf die Körpermasse einer Trakehner Zuchtpferdepopulation. *Züchtungskunde,* 63, 335 – 341.

Kearsley, C. (2007). Genetic Evaluation of Sport Horses in Britain. PhD Thesis. The University of Edinburgh, UK.

Kearsley, C.G.S., Woolliams, J.A., Coffey, M.P. & Brotherstone, S. (2008). Use of competition data for genetic evaluations of eventing horses in Britain: Analysis of the dressage, showjumping and cross country phases of eventing competition. *Livestock Science,* 118, 72–81.

Koenen, E., (2002). Genetic evaluations for competition traits of warmblood sport horses. *WBFSH seminar, Budapest*, 5th November.

Koenen, E.P.C. & Aldridge, L.I. (2002). Testing and genetic evaluation of sport horses in an international perspective. *7th World Congress on Genetics Applied to Livestock Production, Montpellier*, 19th – 23rd August.

Koenen, E.P.C., Aldridge, L.I. & Philipsson, J. (2004). An overview of breeding objectives for warmblood sport horses. *Livestock Production Science,* 88, 77–84.

Knol, D.L. & Ten Berge, J.M.F. (1989). Least-squares approximation of an improper correlation matrix by a proper one. *Psychometrika,* 54, 53–61.

Lande, R. & Arnold, S.J. (1983). The measurement of selection on correlated characters. *Evolution,* 37, 1210 – 1226.

Langlois, B. (1975). Analyse statistique et génétique des gains des Pur Sang anglais de trois ans dans les courses plates francaises. *Annales de Génétique et de Sélection Animale,* 7, 387-408.

Langlois B. (1980). Estimation de la valeur génétique des chevaux de sport d'après les sommes gagnées dans les compétitions équestres françaises. *Annales de Génétique et de Sélection Animale,* 12, 15-31.

Lüehrs-Behnke, H., Röehe, R. & Kalm, E. (2002). Genetic associations among traits of the new integrated breeding evaluation method used for selection of German warmblood horses. *Veterinariay Zootecnia Journal*, 18, 90 – 93.

Lüehrs-Behnke, H., Röehe, R. & Kalm, E. (2006a). Genetische Analysen von Reitsportprüfungen und deren Beziehungen zu Merkmalen der Hengstleistungs- und Zuchtstutenprüfung. *Züchtungskunde,* 78, S. 119 – 128.

Lüehrs-Behnke, H., Röehe, R. & Kalm, E. (2006b). Genetische Parameter für Zuchtstutenprüfungsmerkmale der verschiedenen deutschen Warmblutzuchtverbände. *Züchtungskunde,* 78, S. 271 – 280.

Lüehrs-Behnke, H., Röehe, R. & Kalm, E. (2006c). Schätzung genetischer Parameter für Turniersportleistungen innerhalb Prüfungsklassen und deren Beziehungen zu den Merkmalen der Zuchtstuten- und Hengstleistungsprüfung. *Züchtungskunde,* 78, S. 173 – 183.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates Inc., Sunderland MA, USA.

Mahon, G.A.T. & Cunningham, E.P.C. (1982). Inbreeding and the inheritance of fertility in the thoroughbred mare. *Livestock Production Science,* 9, 743–754.

McMahon, H. (1996). *Leading Competition Sires of GB & Ireland 1996/97*. Otterswick Publishing.

Mantysaari, E.A. (2004). Multiple-trait across-country evaluations using singular (co)variance matrix and random regression model. *Proceedings of the 2004 Interbull Meeting, Sousse, Tunisia.*29 – 31 May.

Meuwissen, T.H., Hayes, B.J. & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics,* 157, 1819–1829.

Meyer, K., Johnston, D.J. & Graser, H.U. (2004). Estimates of the complete genetic covariance matrix for traits in multi-trait genetic evaluation of Australian Hereford cattle. *Australian Journal of Agricultural Research*, 55, 195 – 210.

Miglior, F., Pagnacco, G. & Samore, A.B. (1998). A total merit index for the Italian Haflinger horse using breeding values predicted by a multi-trait animal model. *Proceedings of the 6th World congress on genetics applied to Livestock Production, Armidale, NSW, Australia*, 11th–16th January.

Misztal, I. (2008). Reliable computing in estimation of variance components. *Journal of Animal Breeding and Genetics,* 125, 363–370.

Mrode, R.A. (2005). *Linear Models for the Prediction of Animal Breeding Values*, 2nd edition. CABI Publishing, Wallingford, UK.

Oki, H., Kusunose, R., Nakaoka, H., Nishiura, A., Miyake, T. & Sasaki, Y. (2007). Estimation of heritability and genetic correlation for behavioural responses by Gibbs sampling in the Thoroughbred racehorse. *Journal of Animal Breeding and Genetics,* 124, 185–191.

Olsson, E. (2006). Multi-trait Evaluation of Swedish Warmblood Stallions at Station Performance Tests including Field and Competition Records. Licenticate thesis. Publication no. 144. The Swedish University of Agricultural Sciences, Sweden.

Olsson, E.G., Árnason, T., Nasholm, A. & Philipsson, J. (2000). Genetic parameters for traits at performance test of stallions and correlations with traits at progeny tests in Swedish warmblood horses. *Livestock Production Science,* 65, 81–89.

Olsson, E., Näsholm, A., Strandberg, E. & Philipsson, J. (2008). Use of field records and competition results in genetic evaluation of station performance tested Swedish warmblood stallions. *Livestock Science,* 117, 287–297.

Patterson, H.D & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika,* 58, 545–554.

Philipsson, J. (2007). Considerations for the definition of Warmblood Horses in Canada. [online]. Available from: http://www.canadianwarmbloods.com/2007news/pdfdocs07/ConsiderationsDefinition-ofWarmbloods.pdf. [Accessed May 2011]

Philipsson, J., Árnason, T. & Bergsten, K. (1990). Alternative Selection-Strategies for Performance of the Swedish Warmblood Horse. *Livestock Production Science,* 24, 273–285.

Phillips, P.C. & Arnold, S.J. (1989) Visualising multivariate selection. *Evolution* 43, 1209-1222.

Phocas, F. & Laloë, D. (2004). Should genetic groups be fitted in BLUP evaluation? Practical answer for the French AI beef sire evaluation. *Genetics Selection and Evolution,* 36, 325–345.

Powell, R.L. & Norman, H.D. (2006). Major advances in genetic evaluation techniques. *Journal of Dairy Science* 89, 1337-1348.

Preisinger, R., Wilkens, J. & Kalm, E. (1991). Estimation of genetic parameters and breeding values for conformation traits for foals and mares in the Trakehner population and their practical implication. *Livestock Production Science,* 29, 77-86.

Quinn, (2005). Genetic Evaluation of Show Jumping horses in Ireland. Irish Horse Board. Maynooth, Co Kildare

Reilly, M., Foran, M.K., Kelleher, D.L., Flanagan, M.J. & Brophy, P.O. (1998). Estimation of genetic value of show jumping horses from the ranking of all performances in all competitions. *Journal of Animal Breeding and Genetics,* 115, 17–25.

Ricard, A. (1998). Developments in the genetic evaluation of performance traits in horses. *Proceedings of the 6th World congress on genetics applied to Livestock Production, Armidale, NSW, Australia*, 11–16 January.

Ricard, A. (2004). Heritability of jumping ability and height of pony breeds in France. *Livestock Production Science,* 89, 243-251.

Ricard, A. & Chanu, I. (2001). Genetic parameters of eventing horse competition in France. *Genetics Selection and Evolution,* 33, 175–190.

Ricard, A., Bruns, E. & Cunningham, E.P. (2000) Genetics of performance traits. In: Bowling AT, Ruvinsky A (eds.) *The genetics of the horse*. CABI Publishing, Wallingford, UK, 411–438.

Ricard, A., Fournet Hanocq, F. (1997). Analysis of factors affecting length of competitive life of jumping horses. *Genetics Selection and Evolution,* 29, 251–267.

Roff D (2000). The evolution of the G matrix: selection or drift? *Heredity* 84, 135-142.

Royston, J.P. (1982). Expected normal-order statistics (exact and approximate). *Journal of the Royal Statistical Society. Series C,* 31, 161–165.

Ruhlmann, C., Bruns, E., Fraehr, E., Philipsson, J., Janssens, S., Quinn, K., Thorén Hellsten, E. & Ricard, A. (2009a). Genetic connectedness between seven European countries for performance in jumping competitions of warmblood riding horses. *Livestock Science,* 120, 75–86.

Ruhlmann, C., Janssens, S., Philipsson, J., Thorén Hellsten, E., Crolly, H., Quinn, K., Manfredi, E. & Ricard, A. (2009b). Genetic correlations between horse show jumping competition traits in five European countries. *Livestock Science,* 122, 234–240.

Saastamoinen, M.T. & Barrey, E. (2000). Genetics of conformation, locomotion and physiological traits, in: Bowling, A.T., Ruvinsky, A., (Eds.), *The Genetics of the Horse*. CABI Publishing, Oxon/New York, p.439 – 472.

Sanders, K., Bennewitz, J. & Kalm, E. (2006). Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. *Journal of Dairy Science,* 89, 315–321.

Seidlitz, G., Willeke, H. & von Butler-Wemken, I. (1991). Korpermasse und Exterieurbeurteilungen bei Zuchtstuten des Arabischen Vollblutpferdes. *Archiv Fuer Tierzucht,* 34, 233 – 240.

Sigurdsson, A., Hugason, K. & Árnason, T. (1997). Breeding strategies and genetic progress in the Icelandic toelter population. *48th Annual Meeting of the European Association of Animal Production, Vienna, Austria*. 25th – 28th August.

Stewart, I.D., Woolliams, J.A. & Brotherstone, S. (2010). Genetic evaluation of horses for performance in dressage competitions in Great Britain. *Livestock Science,* 128, 36–45.

Stewart, I.D., Brotherstone, S., White, I.M.S. & Woolliams, J.A. (in submission a). Predicting variance components and breeding values for eventing disciplines and grades in sport horses.

Stewart, I.D., Woolliams, J.A. & Brotherstone, S. (in submission b) Genetic evaluations of traits recorded in British young horse tests.

Stewart, I.D. (2011). Genetic Evaluation of Traits Recorded in Sport Horses in GB. PhD Thesis. The University of Edinburgh.

Stock, K.F. & Distl, O. (2005). Prediction of breeding values for osseous fragments in fetlock and hock joints, deforming arthropathy in hock joints, and pathologic changes in the navicular bones of Hanoverian Warmblood horses. *Livestock Production Science,* 92, 77–94.

Stock, K.F. & Distl, O. (2007). Genetic correlations between performance traits and radiographic findings in the limbs of German Warmblood riding horses. *Journal of Animal Science,* 85, 31–41.

Stock, K.F. & Distl, O. (2008). Multiple-trait selection for radiographic health of the limbs, conformation and performance in Warmblood riding horses. *Animal,* 2, 1724–1732.

Stock, K.F., Hoeschele, I. & Distl, O. (2007). Estimation of genetic parameters and prediction of breeding values for multivariate threshold and continuous data in a simulated horse population using Gibbs sampling and residual maximum likelihood. *Journal of Animal Breeding and Genetics,* 124, 308–319.

Tavernier, A. (1990). Estimation of breeding value of jumping horses from their ranks. *Livestock Production Science,* 26, 277–290.

The Top Dressage Sires of International Horses 1990-1995 (2004). [online]. Available from: http://www.sporthorse-breeder.com/cgi-bin/csArticles/articles/000005/000532.htm. [Accessed July 2011]

Thorén Hellsten, E., Jorjani, H. & Philipsson, J. (2008). Connectedness among five European sport horse populations. *Livestock Science,* 118, 147–156.

Thorén Hellsten, E., Jorjani, H. & Philipsson, J. (2009a). Genetic correlations between similar traits in the Danish and Swedish Warmblood sport horse populations. *Livestock Science,* 124, 15–20.

Thorén Hellsten, E., Nasholm, A., Jorjani, H., Strandberg, E. & Philipsson, J. (2009b). Influence of foreign stallions on the Swedish Warmblood breed and its genetic evaluation. *Livestock Science,* 121, 207–214.

Thorén Hellsten, E., Viklund, Å., Koenen, E.P.C., Ricard, A., Bruns, E. & Philipsson, J. (2006). Review of genetic parameters estimated at stallion and young horse performance tests and their correlations with later results in dressage and show-jumping competition. *Livestock Science,* 103, 1–12.

Tozaki, T., Miyake, T., Kakoi, H., Gawahara, H., Sugita, S., Hasegawa, T., Ishida, N., Hirota, K. & Nakano, Y. (2010). A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. *Animal Genetics,* 41, 28–35.

van Bergen, H.M.J.M. & van Arendonk, J.A.M., (1993). Genetic parameters for linear type traits in Shetland Ponies. *Livestock Production Science,* 36, 273 – 284.

Valberg, S.J. (1997) Polysaccharide storage myopathy (PSSM) *Proceedings of the 15th Annual Veterinary Medical Forum*. American College of Veterinary Medicine, Colorado.

Viklund, Å., Thorén Hellsten, E., Näsholm, A., Strandberg, E. & Philipsson, J. (2008) Genetic parameters for traits evaluated at field tests of 3- and 4-year-old Swedish Warmblood horses. *Animal,* 2, 1832–1841.

Viklund, Å., Braam, Å., Näsholm, A., Strandberg, E. & Philipsson, J. (2010) Genetic variation in competition traits at different ages and time periods and correlations with traits at field tests of 4-year-old Swedish Warmblood horses. *Animal,* 4, 682–691.

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M.,Gnerre, S., Imsland, F., Lear, T. L., Adelson, D. L., Bailey, E., Bellone, R. R., Bloecker, H., Distl, O., Edgar, R. C., Garber, M., Leeb, T., Mauceli, E., MacLeod, J. N., Penedo, M. C. T., Raison, J. M., Sharpe, T., Vogel, J., Andersson, L., Antczak, D. F., Biagi, T., Binns, M. M., Chowdhary, B. P., Coleman, S. J., Della Valle, G., Fryc, S., Guerin, G., Hasegawa, T., Hill, E. W., Jurka, J., Kiialainen, A., Lindgren, G., Liu, J., Magnani, E., Mickelson, J. R., Murray, J., Nergadze, S. G., Onofrio, R., Pedroni, S., Piras, M. F., Raudsepp, T., Rocchi, M., Roed, K. H., Ryder, O. A., Searle, S., Skow, L., Swinburne, J. E., Syvanen, A. C., Tozaki, T., Valberg, S. J., Vaudin, M., White, J. R., Zody, M. C., Lander, E. S., Lindblad-Toh, K. (2009). Genome sequence, comparative analysis and population genetics of the domestic horse (Equus caballus). *Science,* 326, 865–7.

Wall, E., White, I.M.S., Coffey, M.P. & Brotherstone, S. (2005). The relationship between fertility, rump angle, and selected type information in Holstein-Friesian cows. *Journal of Dairy Science,* 88, 1521–1528.

Wallin, L., Strandberg, E. & Philipsson, J. (2001). Phenotypic relationship between test results of Swedish Warmblood horses as 4-year-olds and longevity. *Livestock Production Science,* 68, 97–105.

Wallin, L., Strandberg, E. & Philipsson, J. (2003). Genetic correlations between field test results of Swedish Warmblood Riding Horses as 4-year-olds and lifetime performance results in dressage and show jumping. *Livestock Production Science,* 82, 61–71.

Wallin, L., Strandberg, E., Philipsson, J. & Göran Dalin (2000). Estimates of longevity and causes of culling and death in Swedish warmblood and coldblood horses. *Livestock Production Science,* 63, 275–289.

Walsh, B. & Blows, M.W. (2009). Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annual Review of Ecology, Evolution and Systematics,* 40, 41-59.

Weatherbys (n.d.), Weatherbys historical overview [online]. Available from: http://www.weatherbys.co.uk/historical_overview. [Accessed May 2011]

Weigel, K., Rekaya, R., Fikse, F., Zwald, N. & Gianola, D. (2000). Data structure and connectedness issues in international dairy sire evaluations. *Proceedings of the 2000 Interbull Meeting. Bled, Slovenia*, May 14–15, 2000. *Interbull Bullentin*, 25, 26–30.

White, I.M.S., Thompson, R. & Brotherstone, S. (1999) Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science,* 82, 632–638.

Woolliams, J.A. (2006). Technical Note: A note on the differential impact of wrong and missing sire information on reliability and gain. *Journal of Dairy Science,* 89, 4901–4902.

Wray, N.R. & Goddard, M.E. (1994). Increasing Long-Term Response to Selection. *Genetics Selection and Evolution,* 26, 431–451.

Further Acknowledgements

This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). The ECDF is partially supported by the eDIKT initiative (http://www.edikt.org.uk).

List of Appendices

Guidelines for routine implementation of breeding values for eventing by the British Equestrian Federation

July 2011

I. Stewart

Contents:

*to be included in version supplied to BEF

Introduction

This document gives recommendations and guidelines on routine estimation of breeding values (EBVs) for sires of eventing horses competing in Great Britain. The evaluation system uses eventing competition data, as well as pedigree information. Previous research has estimated the parameters required to drive routine evaluations. This document details the further requirements for processing of competition data, including suggested software packages for predicting breeding values and gives recommendations on the presentation of EBVs.

Data used in previous work to estimate breeding values

Parameters required for the estimation of breeding values (variance components) have been predicted, and breeding values estimated using a dataset of eventing competition data from British Eventing (BE) covering the period from 13/03/1981 to 14/12/2008. The start of this period was set due to the availability of consistent rider information from this time. This data was made available to researchers at The University of Edinburgh, in the form of three tables of data (in an Access database), titled “Horses”, “ImportedResults” and “ResultDetails”.

Currently, EBVs have been predicted for sires with >= 30 progeny in prenovice or novice grades of competition or with any progeny in intermediate or advanced grades. Given the variance components, routine prediction of EBVs can be performed for all sires in the dataset.

In order to predict updated breeding values, various files may be required. These include:

Previous dataset: Dataset_BE_070711_longformat
Sirecode-names: sirecodes_names070711
Gradings (Advanced etc…) assigned for class codes:
“Class_classifications150609_BE_bydiscipline_170609”.
Rider code matchings: Rider table

These can be supplied as easily accessible tables in an Access database.

1. Data Processing

The recommendation is that the datafile that has been prepared covering competitions up to 14/12/2008 is used, and that data from now onwards is prepared and added to the existing dataset. A consistent process of manipulation and formatting must be used, and coded variables must be consistent with the earlier dataset.

No data was actually changed/ deleted. It was manipulated so that new variables are added, and it is always possible to trace back to the original entries.

Data validation - 1st step

1. Horse information: Using data from table "Horses", identify horses based on horse name and competition id. Select relevant horse information: horse id, name, sex, height, heightcm, colour, year foaled, country of origin, dam, sire, current grade and breed. BSJA "horse ids" were retained as the horse identifier.

2. Performance measure:

2.1 Extracting performance data

From table "ResultDetails", extract the following information. Ensure that all records are unique.
Id (horse id), Classref, SectionRef, ResultRef, Position, CompletionCode, EliminationCode, PenaltyCode, Penaltydressagenumeric, SJ, SJTime, XCTime, XCJumping, Discipline, Total

2.2 Cleaning performance data

Clean the performance data, as described in the sections below:

1. exclude records where all results are invalid
2. code eliminations within individual phases of SJ / XC
3. code nulls within individual phases of SJ / XC
4. ensure validity of penalty point data entries – look at values, particularly in extremes, and treat as appropriate
5. remove un-coded eliminations in cross-country / showjumping

2.2.1 Excluding records where all results are invalid

Eventually, the data is to be considered as separate variables for dressage, showjumping and cross-country. This means that when there is a valid result for any phase, the record should be retained, which will occur when a horse finished the dressage, but was eliminated in the next or last phase. The results for the phases in which the horse was eliminated/ did not compete will not be included in the final dataset.

However, it was assumed that if a horse has no dressage penalty points, it was eliminated for all phases. Exclude records where the horse has no dressage penalty points, by

excluding records with dressage penalties null or 0 or with eliminations in the dressage phase. Assume that records with 0 were in error, and represented a null / invalid value; if there is a record with a known valid 0 for dressage penalty points, include this record.

Exclude any records with the "CompletionCode" recorded as "HC" (hors concours), "NR" (non-runner) or "NS" (no show). Hors concours are excluded as the horses competing are not counted in the results of the competition and including them could affect the results of other animals.

2.2.2 Code eliminations and nulls within individual variables by phase

Use the variables: "Penaltydressagenumeric" (dressage penalties), "SJ" (showjumping, jumping penalties), "SJTime" (showjumping, time penalties), "XCTime" (cross-country time penalties), "XCJumping" (cross-country jumping penalties). The variables "Sc time", "sc jumping", "rtb" and "rta", "discipline" and "total penalty points" are not used in the analysis.

Code eliminations (as recorded within the penalty variable for dressage, showjumping, showjumping time, cross-country or cross-country time) as numeric values 1000 or over, to distinguish them from possible valid penalty point values or nulls and to ensure they are numeric. In previous data, these elimination codes within the penalty point variables included the following: (R = retired, E = eliminated, NS = not started, D = disqualified, W = withdrawn).

Code nulls as zero for sj time, sj jumping, xc time and xc jumping.

2.2.3 Clean penalty points data

In general*, penalty points are awarded as follows:
i) for dressage, the points awarded within a test are converted to a percentage, (may be scaled) and are subtracted from 100, so that the better penalty point scores are closer to 0.
ii) SJ jumping penalties are awarded in increments of 4.0, and time in increments of 1.0.
iii) XC jumping penalties are awarded in increments of 20.0, and time 0.4.
* This may not be true for all types of competition/ recording

In order to identify and correct/ exclude any errors, look at the distributions of penalty points in each of the phases individually, to look for inaccuracies in recording. This could be done by producing tables of penalty point values ordered by magnitude, with counts of the numbers of records with each value (a histogram, or other graphical measure could also be produced). Look at those at the extremes, and judge whether these are errors. If obvious errors are found with obvious solutions then correct these. If obvious errors are identified with no obvious solution, then make these values null (i.e. code as >=1000). This takes judgement – we do not want to exclude/ change data incorrectly. No upper limit was imposed on the penalty point values. The overall total recorded within the data "Total" may help with judgements at this point.

Total SJ and XC time and jumping to give an overall score for SJ and one for XC.

NB, in the future, the nature of the recordings by BE may change. This means that some of the data recording – e.g. elimination codings or completion codes may change and any changes will have to be accounted for. Examples of present codings are given on the BE results website, under the results of individual meetings: for example http://www.britisheventing.com/asp-net/Events/Results.aspx?MeetingID=777#completioncodes

2.2.4 Remove un-coded eliminations in cross-country / showjumping

If a horse has no penalty points in SJ or XC phases, this was assumed to be 0 penalty points, except if it was known to have been eliminated in the previous phase of SJ or XC, which depends on the class of competition.

SJ is performed prior to XC in event codes 1 & 2, thus in event codes 1 & 2, if a horse was eliminated in SJ, it was also eliminated in XC. This will be indicated by having event codes 1 or 2, XC total being 0, a dressage total and SJ total being eliminated.

XC is performed before SJ in event codes other than 1 and 2 (for example 3, BC, CCI, CCIO, CCN, CH, CIC, SC) thus if a horse is eliminated in XC, it is also eliminated in SJ. This will be indicated by records having event codes other than 1 and 2, SJ total being 0, a dressage total, and XC total being eliminated.

Identify these records, and recode these 0 values as eliminations in the phases appropriately (i.e. as =>1000).

2.3 Attaching horse and performance data

Attach cleaned performance data from 2 to horse details data from 1, using horse id

2.4 Cleaning height and age

For dataset in 3 clean height, class and age, and add these to the dataset. Ensure records are unique, remove duplicates.

Height is specified in hands/ inches and cm in the original data. Use the height cm value in the first instance (as this should be most accurate), and the hands specification otherwise. Clean any invalid entries, by correcting obvious errors, and setting to missing when there is no obvious solution. Transform height to the nearest cm, potentially by creating a new table that gives conversion values for the entries in the original data. Specify in cm with no decimal places. Check the validity of anomalous values (change or set to null if unknown).

Calculate age, for each record, (in years and fractions of years, to an accuracy of 2 decimal places) based on date of competition and year foaled. Date of birth is assumed to be 1st of January, unless otherwise specified. Check the validity of values, deal with anomalous values appropriately, set to null if unknown. Later in the system, records where the age is <4 years will be excluded and records =>20 years will be grouped as 20.

2.5 Including record id
Take this dataset, and add a record id (consistent with, and continuing on from the final record id in the previous file). Exclude duplicate records.

2.6 Including class classifications
Work out class classifications (Prenovice, Novice, Intermediate or Advanced; coded as P, N, I or A respectively) for each phase individually and as a total competition classification based on classcode. For prior classifications, see Table in appendix 1. "Class_classifications150609_BE_bydiscipline_170609".

Take dataset in 5, restrict records to only those with a total classification A, I, N or P and add classifications for each of the individual disciplines.

2.7 Replace eliminations with nulls
Take data in 6 and replace eliminations (=>1000) with nulls

For information, this dataset includes the following variables:
Record id, horse id, horse name, sex, age (yrs), height (cm), dressage classification, showjumping classification, cross-country classification, total classification, BE section ref, BE class ref, BE rider member ref, sire (name), dam (name), dressage penalty points, showjumping penalty points, cross-country penalty points.

Transformation of penalty points to a normal score

From 7 extract performance data, in the form of penalty points for each discipline to calculate normal scores. Normal scores must be calculated based on a dataset that is as complete as possible, in order to have accurate rankings within the competitive class (Classref). Therefore, transformation is performed prior to exclusion of records with missing model variables, (horse gender, age, class, sire and rider) from the dataset to ensure that the number of competitors in each class is as complete as possible.

The normal score is calculated based on the performance ranking within the class (classrefs), where class is defined as those horses that competed in the same class at the same event.

Extract the following variables: record id, class ref, penalty points for dressage, showjumping and cross-country.

E.g. Performance data for Ranking and Normal scores

Record id	BE Class Reference	dressage	sj	xc
140	303	45	12	120

There are various programs available which calculate normal scores. Fortran routines from the Numerical Algorithms Group (NAG) allow the calculation of both exact and approximate normal scores, depending on the routine used. The g01dhf routine (http://www.nag.co.uk/numeric/Fl/manual/pdf/G01/g01dhf.pdf; NAG Fortran Library 21, Numerical Algorithms Group Limited, UK, 2006) is called by a Fortran program and is one of the few routines which produces exact normal scores. Further information on NAG routines is given in the NAG Fortran Library Manual (NAG Ltd, 2006).
Other software packages that allow the calculation of approximate normal scores, include R (R Development Core Team, 2005) and in the C programming language. If Fortran / NAG is not available we suggest one of the alternative (approximate) routines is used.

For the previous work penalty points were converted to normal scores using the NAG Fortran library routine g01dhf (http://www.nag.co.uk/numeric/Fl/manual/pdf/G01/g01dhf.pdf), and exact values were calculated. The Fortran coding used is given in Appendix 2.

Normal scores were defined to 4 decimal places.

NB in the transformation that was used for the current data, and thus required for future updates, normal scores were calculated so that a **good** performance (corresponding with few penalty points) was converted to a **negative** normal score and a bad performance (corresponding to many penalty points) was converted to a positive normal score.

Data validation - 2nd step

2.8 Coding variables

The data for sex, class and rider should now be coded.

Much of the coding in the previous work was performed in order to ensure data anonymity. Anonymity may not be required when further work is performed by the BEF, however, if the previous dataset is to be used, and simply added to, then coding, manipulation and formatting must be consistent. Levels already in the dataset must be coded as previously performed and new values must be assigned new, unique codes.

Gender was coded as follows: stallion = 1, gelding = 2, mare = 3
Class refs must be coded continuing on from previous maximum (numeric).
Rider references must be coded consistently with previous data, so that riders that already exist are assigned the same code, and new riders are coded with unique codes (see "Coded_riderMemberRef1", which contains BE rider reference numbers and codes that were assigned for the previous analysis). Ensure that any missing riders / errors are coded as null.

Take the dataset in 7, and include the coded variables in the dataset.

The dataset now contains the following variables:
Record id, horse id, gender (coded), age (yrs), height (cm), dressage classification, showjumping classification, cross-country classification, total classification, class code, rider code, sire, dam, dressage penalty points, showjumping penalty points, cross-country penalty points. The previous dataset also included section code; however, this was not used in the analysis, and thus can be coded as a null dummy variable.

2.9 Pedigree validation
Assemble the pedigree. In the previous work, only sire data was of sufficient quality for the analysis. Dam data was insufficiently recorded to identify animals uniquely.

Sires assigned for each horse in the current dataset are given in the dataset, using codes. Sirecodes-names are described in sirecodes_names070711

In the raw data sire names are not recorded consistently. Therefore there may be many different entries for the same sire, due to differences in spelling, and in the presence/ absence of prefixes and suffixes. The numeric coding must identify each sire uniquely, with a single code. Each numeric sire code assigned may then refer to multiple versions of one name. The current sire codes must be used where the sire already exists in the data. Code any new sires using a new and unique id. In the previous work, considered research interpretation was required for assignments. Update any old assignments if deemed appropriate.

Pedigree Data – general information
The previous work predicted EBVs for sires, using a "sire model". The pedigree information available (obtained in the dataset from BE) was sufficient for the identification of sires of competing horses, but not for dams or any further pedigree. Potentially, in the future, breeding values could be predicted for more animals in the pedigree; this will require better pedigree information, and would use an "animal model". A separate file containing pedigree information would be required for this analysis, containing details of the parents of competing horses, grandparents, and as many further generations as possible, with the same horse consistently identified across generations.

2.10 Attach normal scores
Take dataset in 8 and attach normal scores (using record id) and sires (using horse id), for each record.

2.11 Data check
Final check over data to ensure that all is correct.

Exclude age values <4 and replace values greater than 20 with 20. NB values <4 and >20 are included in the last dataset, and were excluded/ grouped at the point of analysis, so these may need to be altered also, unless this can be performed at the point of analysis.

2.12 Exclude records with missing variables

Exclude records with missing variables that are required for the model, i.e. exclude records with missing gender, age, class, sire, and rider.

2.13 Formatting dataset

Ensure that variables are in a consistent format with the original dataset (numeric etc….)

Data may require re-formatting in order to make it consistent with the BLUP program chosen. To be consistent with the current dataset, the formatting required will be as illustrated in table 1.

Table 1. Sample of final dataset

ID1	id	sirerecode070711	sex_coded	age_yrs2	age_yrs2_nodp	Heightcm_nodp	DressageClassification	SJClassification	XCClassification	TotalClassification	Section	Class	Rider	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
51	22	2701	3	9.5	10	163	P	P	P	P	5689	3210	4698				2.2170				-0.1710				0.5010
166	81	9593	2	10.6	11	170	I	I	I	I	5068	2974	3555		1.2820				-0.3230				0.8940		
166	81	9593	2	10.6	11	170	I	I	I	I	5068	2974	3555		1.2820				-0.3230				0.8940		
166	81	9593	2	10.6	11	170	I	I	I	I	5068	2974	3555		1.2820				-0.3230				0.8940		
167	81	9593	2	10.6	11	170	I	I	I	I	5077	2978	3555		1.0900				-0.6910				-0.0370		
167	81	9593	2	10.6	11	170	I	I	I	I	5077	2978	3555		1.0900				-0.6910				-0.0370		
167	81	9593	2	10.6	11	170	I	I	I	I	5077	2978	3555		1.0900				-0.6910				-0.0370		
168	81	9593	2	10.6	11	170	N	N	N	N	5090	2986	3555			2.0358				0.6297				-0.1880	
168	81	9593	2	10.6	11	170	N	N	N	N	5090	2986	3555			2.0358				0.6297				-0.1880	
168	81	9593	2	10.6	11	170	N	N	N	N	5090	2986	3555			2.0358				0.6297				-0.1880	
169	81	9593	2	11.4	11	170	N	N	N	N	5417	3111	5628			2.5134				1.5861					
169	81	9593	2	11.4	11	170	N	N	N	N	5417	3111	5628			2.5134				1.5861					
169	81	9593	2	11.4	11	170	N	N	N	N	5417	3111	5628			2.5134				1.5861					
170	81	9593	2	12.4	12	170	N	N	N	N	6310	3504	5178			1.1609				-0.8530				0.5677	
170	81	9593	2	12.4	12	170	N	N	N	N	6310	3504	5178			1.1609				-0.8530				0.5677	

Table 2. Descriptions of variables in dataset

Variable label	Description	notes
ID1	record id	
id	horse id	
sirecode080909	sire code	
sex_coded	Gender code	
age_yrs2	age (yrs)	Includes all ages, even <4 and >20, In analysis, group these so that <4 are excluded and >= 20 are grouped
Heightcm_nodp	Height (cm)	
DressageClassification	Dressage classification (P, N, I or A)	
SJClassification	Showjumping classification (P, N, I or A)	
XCClassification	Cross-country classification (P, N, I or A)	
TotalClassification	Total classification (P, N, I or A)	
Section	Section code	
Class	Class code	
Rider	Rider code	
DA	Dressage advanced normal score	
DI	Dressage intermediate normal score	
DN	Dressage novice normal score	
DP	Dressage prenovice normal score	
SJA	Showjumping advanced normal score	
SJI	Showjumping intermediate normal score	
SJN	Showjumping novice normal score	
SJP	Showjumping prenovice normal score	
XCA	Cross-country advanced normal score	
XCI	Cross-country intermediate normal score	
XCN	Cross-country novice normal score	
XCP	Cross-country prenovice normal score	

BLUP software alternatives

Re-organisation and formatting of the data file to fit the package may be required.

The previous analysis used ASReml (Gilmour et al., 2006) to estimate the parameters required for BLUP predictions of sire breeding values, and to perform the BLUP analysis. As variance components have now been estimated, these can be used to predict breeding values using alternative packages. There are a number of different packages, of which a few are noted here.

PEST

This package was produced by Groeneveld, E et al., (1990) and is widely used.
It is written in Fortran 77/90. There are no inherent limitations on the number of traits, fixed or random effects or covariables. It is driven by a parameter file and operates on codes of up to 16 characters in length. The manual is available at:
ftp://192.108.34.51/pub/pest/doc/pest-manual-Apr-2006.pdf.
This package only handles reliability calculations for small datasets so an approximate method would need to be employed for the BE data.

The type of license required would be an unrestricted production license, which would cost €2750, as a one-off fee.
For further information contact Eildert Groeneveld (email: eildert.groeneveld@onlinehome.de).
To obtain the software supply name and address of license holder and operating system on which it would be used).

Groeneveld, E.; Kovac, M.; Wang, T. (1990). PEST, a general purpose BLUP package for multivariate prediction and estimation. Proceedings of the 4th World Congress on Genetics applied to Livestock Production, Edinburgh 23-27 July 1990 pp. 488-491

MiX99 / MiXBLUP

These two packages were developed by Linauder and Mantysaari together with Wageningen University. They are aimed at different user requirements, of which MiXBLUP may be most appropriate in this instance.

MiXBLUP
This package puts emphasis on good customer services and user friendliness, and includes user support for software installation and problems. Different licenses are available depending on requirements. A MiXBLUP Small license (which handles up to 1,000,000 equations) should be sufficient for the BE work, and would cost €1.900, per year. A trial version (for small datasets) can be obtained.
Further information can be found at www.mixblup.eu, or obtained from Han Mulder at Wageningen University.

The package includes reliability calculations.

The alternative MiX99 package is aimed at large evaluation centers and focuses on dairy cattle evaluation. It has greater facilities, however, the yearly license fee is much higher (12000€). Manuals are available at www.mtt.fi/BGE/Software/MiX99.

Reliability calculations

Reliabilities provide a measure of the dependability of the EBV, based on the amount of information that was available to make the estimate.

A further consideration in the choice of BLUP software is the calculation of reliabilities. MixBLUP (and Mix99) include reliability calculations. The calculations within PEST would only be appropriate for a small number of animals, and would be insufficient for the BE work.

The previous work calculated approximate reliabilities, based on the number of progeny with records and the number of records per progeny. The method is documented in Appendix 4 and would be an alternative method of calculating reliabilities. However, such a method would require the input of a programmer.

The Genetic Model

The genetic model required for BLUP prediction is given below.

The dependent variable is the normal score for the competition result. Fixed effects are the horse gender, age (as a quadratic polynomial) and class. Random effects are the sire, horse, rider and residual error. A sire model was used (indicating that sire pedigree information only was used to estimate the genetic variance).

In mathematical notation, the basic sire linear mixed effects model can be described as:

$$Y_{jm\phi} = \mu + \alpha_{i(m)} + \sum \gamma_r x_m^r + \beta_j + a_{s(m)} + c_m + \omega_\phi + e_{jm\phi}$$

where $Y_{jm\Phi}$ is the normal score for rider Φ on horse m in competition class j, $i(m)$ is gender of horse m, $s(m)$ is sire of horse m, $a_{s(m)}$ is breeding value of $s(m)$, x_m is age of horse at time of competition, c_m is the effect of permanent environment for horse m, ω_Φ represents the rider and $e_{jm\Phi}$ is the residual error for horse m, competing in competition class j, with rider Φ.

Regression coefficients γ_1 and γ_2 are assumed constant over grades and disciplines, all other parameters are specific to grade and discipline.

Gender, age and class are treated as fixed effects. Age at time of competition is included as a quadratic polynomial.

Variance component estimates

These parameters are required to drive the BLUP predictions. Having been estimated in previous work, they are detailed in Appendix 3.

Publication reliability limits/ presentation

For publication, a lower threshold limit on the reliability of EBVs should be set. These are recommended at 0.3 for dressage and showjumping, and 0.1 for cross-country (where the heritability and thus reliabilities tend to be lower overall).

The recommendation is that the reliability values for each EBV are presented alongside the EBV value using a graphical scheme. A colour grading scheme, ranging from green (indicating a high reliability) to red (indicating a low reliability) could be used, set within a thermometer shape/ image.

Presentation of EBVs

EBVs should be presented as a numerical value, with the accompanying reliability value. These may either be presented as 12 individual breeding values (one for each discipline-grade trait), or for a selection of discipline-grade traits (for instance for advanced grade in each phase) or as an aggregate index. The values for the different disciplines could potentially be weighted in different ways and then summed to produce a single index figure; alternatively, equal weight could be given to each discipline.
Following the custom in other countries, we recommend that EBVs are transformed prior to presentation. Raw values are transformed to fit a normal (bell shaped) distribution, with a mean of 100 and a standard deviation of 20. This will simply require transformation using the equation given below:

$$\frac{[(uEBV - \mu) \times 20]}{\sqrt{\sigma^2_A}} + 100$$

where uEBV is the unscaled EBV, μ is the total mean of the EBVs and σ^2_A is the genetic variance. The genetic variance for each of the 12 traits is provided in Appendix 5.

Quality assurance

Quality assurance is required throughout the data processing and programming.

As the intention is that the EBVs will be published, it is also necessary that some external consultation is used. The aim of this is to ensure that the procedure of EBV estimation is valid, and that the analysis makes sense theoretically. This would have to be performed by someone with expertise in the area.

An alternative is that the work could be commissioned out. This would likely require someone with suitable expertise to be employed by the BEF for a few weeks each year. Some research groups that perform similar work, and that may be interested in taking the UK BE evaluations on include:

1) A Nordic group who perform routine evaluations for various traits in dairy cattle. Further information on the group is available at:

http://www.nordicebv.info/News/Joint+Nordic+estimation+of+breeding+values+for+yield+traits.htm

2) Steven Janssens or colleagues based at Katholieke Universiteit Leuven, Belgium. Further information available at:
http://www.biw.kuleuven.be/genlog/livgen/chgs_start.html;
http://www.kuleuven.be/wieiswie/nl/person/00015116

3) The Irish Horse Board, Horse Sport Ireland, based in Naas, Co. Kildare.

4) A post-doc researcher with programming experience and suitable expertise in the field

5) Edinburgh Genetic Evaluation Services (EGENES). Further information available at http://www.sac.ac.uk/research/groups/sls/teams/edinburghgeneticevaluations/, although the cost of this is likely to be prohibitive

Updating

It is recommended that breeding values are estimated annually, using updated competition and pedigree information.

Ideally variance components should be re-estimated every few years in case population parameters have changed. This will however, require a more extensive analysis and greater expertise.

References

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2006). ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. (2009). ASReml User Guide Release 3.0 VSN International Ltd, Hemel Hempstead, UK www.vsni.co.uk.

NAG Ltd, *NAG Fortran Library Manual, Mark 21*, The Numerical Algorithms Group Limited, UK, 2006.

R Development Core Team (2005). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing
Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

List of Appendices

Appendix 1. Grade classifications (Advanced, Intermediate, Novice or Prenovice) assigned for different class codes

A = Advanced, I = Intermediate, N = Novice, P = Prenovice

Class Code	Description	Dressage Classification	Showjumping Classification	Cross-country Classification	TotalClassification
*	international novice	N	N	N	N
**	international intermediate	I	I	I	I
***	international advanced	A	A	A	A
****	international championship	A	A	A	A
***W	world cup qualifier	A	A	A	A
***WF	world cup final	A	A	A	A
4YO		Intro	intro	intro	intro
5YO		P	P	P	P
6YO		N	N	N	N
A	advanced	A	A	A	A
AI	advanced intermediate	I	I	I	I
AU25	advanced under 25	A	A	A	A
I	intermediate	I	I	I	I
IN	intermediate novice	I	N	N	N
Intro	Intro (Grade 4 only)	P	P	P	P
Intro T&C		Intro	Intro	Intro	Intro
INU21	intermediate novice under 21	I	N	N	N
IRF	intro regional final	P	P	P	P
J*	junior international novice	N	N	N	N
J**		I	I	I	I
JI	junior intermediate novice	I	I	I	I
JIN	junior intermediate novice	N	N	N	N
JN	junior novice	N	N	N	N
JOI	junior open intermediate	I	I	I	I
JON	juniour open novice	N	N	N	N
JPN	junior pre novice	P	P	P	P
JRN	junior regional novice	N	N	N	N
M					

MN	military novice	N	N	N	N
MON	military open novice	N	N	N	N
MOPN	military open pre novice	P	P	P	P
MPN	military pre novice	P	P	P	P
N	novice	N	N	N	N
NRF	novice regional final	N	N	N	N
O CH	open championships	A	A	A	A
OG***		A	A	A	A
OI	open intermediate	I	I	I	I
OIntro		Intro	Intro	Intro	Intro
OIU21	open intermediate under 21	I	I	I	I
OIU25	open intermediate under 25	I	I	I	I
ON	open novice	N	N	N	N
OPN	open pre novice	P	P	P	P
P	pony	P	P	P	P
P**	pony 2 star	N	N	N	N
PCI	pony club intro	Intro	Intro	Intro	Intro
PCPN	pony club pre novice	P	P	P	P
PN	pre novice	P	P	P	P
PNP		N	P	P	P
PNRF	prenovice regional final	P	P	P	P
PNU21	pre novice u 21	P	P	P	P
PPN	pony prenovice	P	P	P	P
PT		N	N	N	N
RCPN		P	P	P	P
T	training	pre-intro	pre-intro	pre-intro	pre-intro
VPN	veteran pre novice	P	P	P	P
YH*	young horse 1 star	N	N	N	N
YR**	young rider internatioanl intermediate	I	I	I	I
YRA	young rider advanced	A	A	A	A
YRAI	young rider advanced intermediate	I	I	I	I

YRI	young rider intermediate	I	I	I	I
YROI	young rider open intermediate	I	I	I	I

Appendix 2. Fortran routine used to compute normal scores in previous work

This uses NAG fortran library routine g01dhf

(http://www.nag.co.uk/numeric/Fl/manual/pdf/G01/g01dhf.pdf)

Data was first separated into a different file for each of the disciplines (each containing record id, class id, and penalty points). Data was then sorted by competition class.

```
      character*1 :: scores, ties
      integer :: Class (361), ID1(361)
      real :: Dressagepenalties(361)
      integer :: n,iwrk(361),ifail
      double precision :: x(361),r(361)
      open(10,file='Dressage_rawdata030809sort.txt')
      open(11,file='dressagenscore030809.txt')
      iflg=0
      i=1
      scores='N'
      ties='A'
      ifail=0

1     read(10,*,end=99) Class (i), Dressagepenalties(i), ID1(i)
      if(Class(i).ne.Class(1))go to 2
      i=i+1
      go to 1
2     n=i-1

      print *,n,' in class ',Class(1)
      do 3 k=1,n
      x(k)=real(Dressagepenalties(k))
3     continue

      call g01dhf(scores,ties,n,x,r,iwrk,ifail)

      do 4 k=1,n
      write(11,100)ID1(k),Class(k),Dressagepenalties(k),r(k)
4     continue

      ID1(1)=ID1(i)
      Class(1)=Class(i)
      Dressagepenalties(1)=Dressagepenalties(i)
      i=2
      if(iflg.eq.0)go to 1
      go to 98
99    iflg=1
```

```
      go to 2
98    stop
100   format(2i7,i4,f9.4)
      end
```

Appendix 3. (Co)variance, variance, correlation matrices for the twelve traits, for the sire (genetic), rider, horse and residual error

Sire (co)variance, variance, correlation matrix for the twelve traits
(Co)variances between traits are given below the diagonal, variances are given on the diagonal and correlations between traits above the diagonal
The equivalent genetic variances are 4* this value

	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
DA	0.02221	0.91700	0.79100	0.58770	-0.57150	-0.66630	-0.16450	-0.04631	-0.00525	-0.25480	0.25330	-0.33820
DI	0.01915	0.01965	0.98680	0.94150	-0.35590	-0.03971	0.13810	0.12990	-0.64600	-0.64490	0.11400	-0.20880
DN	0.01578	0.01851	0.01791	0.98630	-0.21380	0.13990	0.12360	0.22990	-0.23340	-0.55860	-0.30950	-0.10910
DP	0.01212	0.01826	0.01826	0.01914	-0.06967	0.19690	0.20270	0.31940	-0.63010	-0.08341	-0.49600	0.00591
SJA	-0.01668	-0.00977	-0.00561	-0.00189	0.03837	0.90210	0.89490	0.73950	-0.35830	-0.56180	-0.00206	0.19040
SJI	-0.01459	-0.00082	0.00275	0.00400	0.02597	0.02159	0.98540	0.94810	0.02524	0.00028	0.15430	0.22520
SJN	-0.00349	0.00276	0.00236	0.00399	0.02497	0.02062	0.02029	0.95140	-0.06811	0.07081	0.27330	0.14150
SJP	-0.00094	0.00249	0.00420	0.00603	0.01977	0.01901	0.01849	0.01862	-0.31720	-0.22930	0.20810	0.26030
XCA	-0.00006	-0.00732	-0.00253	-0.00705	-0.00568	0.00030	-0.00078	-0.00350	0.00654	0.98920	0.89660	0.56060
XCI	-0.00105	-0.00250	-0.00207	-0.00032	-0.00304	0.00000	0.00028	-0.00087	0.00221	0.00076	0.66480	0.04583
XCN	0.00214	0.00091	-0.00235	-0.00390	-0.00002	0.00129	0.00221	0.00161	0.00412	0.00104	0.00322	0.75270
XCP	-0.00267	-0.00155	-0.00077	0.00004	0.00198	0.00175	0.00107	0.00188	0.00240	0.00007	0.00226	0.00281

Rider (co)variance, variance, correlation matrix for the twelve traits
(Co)variances between traits are given below the diagonal, variances are given on the diagonal and correlations between traits above the diagonal

	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
DA	0.2875	0.9677	0.9059	0.8067	0.7924	0.7400	0.6061	0.5387	0.6098	0.3552	0.3518	0.3239
DI	0.2607	0.2524	0.9736	0.9152	0.7632	0.7584	0.6646	0.6448	0.5430	0.3869	0.3973	0.4148
DN	0.2287	0.2303	0.2218	0.9723	0.6532	0.7334	0.6006	0.6105	0.5736	0.4626	0.4409	0.4575
DP	0.2094	0.2226	0.2217	0.2344	0.5131	0.6444	0.5340	0.5491	0.4384	0.4263	0.4041	0.4754
SJA	0.1391	0.1256	0.1007	0.0813	0.1072	0.9675	0.8926	0.7901	0.5659	0.4767	0.4677	0.3782
SJI	0.1269	0.1218	0.1104	0.0997	0.1013	0.1022	0.9944	0.9061	0.6135	0.5714	0.5274	0.4290
SJN	0.0909	0.0933	0.0791	0.0723	0.0817	0.0889	0.0781	0.9046	0.6196	0.5440	0.5682	0.4438
SJP	0.0766	0.0859	0.0762	0.0705	0.0686	0.0768	0.0671	0.0703	0.6186	0.6107	0.5480	0.6087
XCA	0.1106	0.0923	0.0914	0.0718	0.0627	0.0664	0.0586	0.0555	0.1144	0.8536	0.7649	0.6265
XCI	0.0645	0.0658	0.0738	0.0699	0.0529	0.0619	0.0515	0.0549	0.0978	0.1147	0.9185	0.6826
XCN	0.0633	0.0670	0.0697	0.0656	0.0514	0.0566	0.0533	0.0487	0.0868	0.1044	0.1125	0.7831
XCP	0.0471	0.0565	0.0584	0.0624	0.0336	0.0372	0.0336	0.0437	0.0574	0.0627	0.0712	0.0735

Horse (co)variance, variance, correlation matrix for the twelve traits
(Co)variances between traits are given below the diagonal, variances are given on the diagonal and correlations between traits above the diagonal

	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
DA	0.2725	0.9688	0.7463	0.6812	0.2024	0.0632	-0.0409	-0.0596	0.1840	0.0294	-0.0780	0.0983
DI	0.2597	0.2638	0.9167	0.7386	0.0935	0.1048	-0.0039	-0.0600	0.1715	0.1696	0.0318	0.0169
DN	0.1833	0.2214	0.2213	0.8959	0.0281	-0.0089	0.1132	0.0677	-0.0273	0.0230	0.1356	0.0744
DP	0.1702	0.1815	0.2017	0.2290	0.0405	-0.0117	0.0503	0.1394	0.0128	-0.0491	0.0520	0.1608
SJA	0.0425	0.0193	0.0053	0.0078	0.1618	0.9629	0.7059	0.3624	0.4072	0.3242	0.1483	0.0536
SJI	0.0120	0.0196	-0.0015	-0.0020	0.1407	0.1320	0.9115	0.6451	0.2354	0.3192	0.0507	-0.0306
SJN	-0.0076	-0.0007	0.0189	0.0085	0.1006	0.1173	0.1255	0.8418	-0.0091	0.1155	0.3388	0.0488
SJP	-0.0102	-0.0101	0.0104	0.0219	0.0478	0.0769	0.0979	0.1077	0.0443	0.0212	0.1183	0.2730
XCA	0.0276	0.0253	-0.0037	0.0018	0.0471	0.0246	-0.0009	0.0042	0.0826	0.8004	0.2644	0.0989
XCI	0.0045	0.0258	0.0032	-0.0070	0.0386	0.0343	0.0121	0.0021	0.0681	0.0875	0.6914	0.2831
XCN	-0.0123	0.0049	0.0193	0.0075	0.0180	0.0056	0.0363	0.0117	0.0230	0.0618	0.0912	0.4930
XCP	0.0131	0.0022	0.0089	0.0196	0.0055	-0.0028	0.0044	0.0229	0.0072	0.0214	0.0380	0.0650

Residual error (co)variance, variance, correlation matrix for the twelve traits
(Co)variances between traits are given below the diagonal, variances are given on the diagonal and correlations between traits above the diagonal

	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
DA	0.4051	0.0000	0.0000	0.0000	0.0191	0.0000	0.0000	0.0000	0.0674	0.0000	0.0000	0.0000
DI	0.0000	0.5006	0.0000	0.0000	0.0000	0.0207	0.0000	0.0000	0.0000	0.0942	0.0000	0.0000
DN	0.0000	0.0000	0.5316	0.0000	0.0000	0.0000	0.0194	0.0000	0.0000	0.0000	0.0811	0.0000
DP	0.0000	0.0000	0.0000	0.5281	0.0000	0.0000	0.0000	0.0266	0.0000	0.0000	0.0000	0.0633
SJA	0.0097	0.0000	0.0000	0.0000	0.6392	0.0000	0.0000	0.0000	0.1480	0.0000	0.0000	0.0000
SJI	0.0000	0.0118	0.0000	0.0000	0.0000	0.6525	0.0000	0.0000	0.0000	0.1621	0.0000	0.0000
SJN	0.0000	0.0000	0.0113	0.0000	0.0000	0.0000	0.6394	0.0000	0.0000	0.0000	0.1513	0.0000
SJP	0.0000	0.0000	0.0000	0.0155	0.0000	0.0000	0.0000	0.6448	0.0000	0.0000	0.0000	0.1284
XCA	0.0372	0.0000	0.0000	0.0000	0.1026	0.0000	0.0000	0.0000	0.7521	0.0000	0.0000	0.0000
XCI	0.0000	0.0586	0.0000	0.0000	0.0000	0.1150	0.0000	0.0000	0.0000	0.7714	0.0000	0.0000
XCN	0.0000	0.0000	0.0513	0.0000	0.0000	0.0000	0.1050	0.0000	0.0000	0.0000	0.7531	0.0000
XCP	0.0000	0.0000	0.0000	0.0373	0.0000	0.0000	0.0000	0.0837	0.0000	0.0000	0.0000	0.6591

Appendix 4. Method used to calculate approximate reliabilities in previous work

Reliabilities were calculated for each sire, for each grade-discipline, based on the number of progeny records, adapting a method published by Harris & Johnson (1998). These are approximate reliabilities, which were used as the standard error estimates of EBVs, as described below, were not provided by the software in this instance. The method was adapted for the current work, as records were available only for the progeny of a sire. The original method also included information from the sire's records, and those of other relatives.

The following equation was used:

$$R(p) = \frac{\bar{n}h^2 m}{4 + (m - 1)\bar{n}h^2 + 4(\bar{n} - 1)r}$$

(Harris & Johnson 1998).

Where R(p) = the reliability based on progeny,
h^2 = the heritability of the trait,
r = the repeatability of the trait,
m = number of progeny with records, and
$\bar{n}$ = the average number of records per progeny adjusted for the competition class. This was calculated for each sire, as the average of n for all its progeny, where n for each horse was calculated by:

$$n = \sum_{i=1}^{1} (1 - \frac{1}{ncg_i})$$

(Harris & Johnson 1998).

Where ncg_i is the number of records in class i.

This method does not account for the fact that the analysis was multivariate, and so will give a lower bound to the reliability (i.e. the reliability will be biased downwards). It does however account for the number of records per sire and the number of progeny.

EBVs were presented with a mean of 100 and a standard deviation of 20, scaled by the genetic standard deviation, using the following equation:

$$\frac{[(uEBV - \mu)*20]}{\sqrt{\sigma^2_A}} + 100$$

Where μ is the mean EBV, uEBV is the unscaled EBV and σ^2_A is the genetic variance.

It has been suggested that when EBVs are published, reliabilities are presented in bands, e.g. <50, 50-65, 66-80, 81-95, >95 (to be decided in consultation with the industry), or that colour coding is used (e.g. red, amber, green), rather than presenting exact figures.

Alternative methods

Kearsley et al. used the standard error (s.e.) for the EBV to calculate the prediction error variance (PEV). Our software (ASReml) usually provides a s.e. for the EBV, however, in the current instance no estimate was produced, due to the complexities of the computation. The s.e. estimate provided is the square root of the prediction error variance (PEV), which is based on the amount of information included in the EBV. The PEV is (1-reliability)*σ^2_A, where σ^2_A is the additive genetic variance.

Alternative methods used in the literature include, Tier & Meyer (2004), Banos et al (2004) and Meyer (1989).

References:

Banos G, Brotherstone S, Thompson R, Woolliams JA, Wall E & Coffey MP (2004). Calculation of multiple-trait sire reliability for traits included in a dairy cattle fertility index. *Animal Science* 79:1-9

Gilmour AR, Gogel BJ, Cullis BR, & Thompson R (2006). ASReml User Guide Release 2.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK

Harris B & Johnson D (1998). Approximate reliability of genetic evaluations under an animal model. *Journal of Dairy Science* 81(10): 2723 – 2728.

Meyer K (1989). Approximate accuracy of genetic evaluation under an animal model. *Livestock Production Science* 21:87-100

Tier B & Meyer K (2004). Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. *Journal of Animal Breeding and Genetics* 121:77–89

Appendix 5. Genetic variance estimates required for transformation of EBVs to publication scale

Trait	Genetic Variance
DA	0.08884
DI	0.07860
DN	0.07164
DP	0.07656
SJA	0.15348
SJI	0.08636
SJN	0.08116
SJP	0.07448
XCA	0.02614
XCI	0.0030576
XCN	0.012896
XCP	0.011224

Appendix 6. Example ASReml .as file – (bivariate analysis used in estimation of variance components)

```
BE analysis
 ID1 !I 345067
 id !I 19829
 sirecoded !I 3017 !D
 sex_coded !I 3 !D
 age_yrs2
 age_yrs2_nodp !D !D<4
 Heightcm_nodp
 DressageClassification !A 4
 SJClassification !A 4
 XCClassification !A 4
 TotalClassification !A 4
 Section !I 15680
 Class !I 6875 !D
 Rider !I 11841 !D
 DP
 DN
 DI
 DA
 SJP
 SJN
 SJI
 SJA
 XCP
 XCN
 XCI
 XCA
 age_yrs3 !=age_yrs2_nodp !REPLACE 26 20
 age_yrs4 !=age_yrs3 !REPLACE 25 20
 age_yrs5 !=age_yrs4 !REPLACE 24 20
 age_yrs6 !=age_yrs5 !REPLACE 23 20
 age_yrs7 !=age_yrs6 !REPLACE 22 20
 age_yrs8 !=age_yrs7 !REPLACE 21 20
 age_yrs9 !=age_yrs8 !REPLACE 27 20
ShowjumpingA_ShowjumpingI_All.csv !Skip 1 !NOD !FCON !BRIEF -1 !CONTINUE !MAXIT 20

SJA SJI ~ Trait Trait.sex_coded Trait.pol(age_yrs9,2) !r Trait.sirecoded Trait.id Trait.Rider !f Trait.Class
 1 2 3
 0
 Trait 0 US 1 0 2 !GPFP
 Trait.sirecoded 2
 Tr 0 US !GP
 3*0
 sirecoded 0 ID !GP
 Trait.id 2
 Tr 0 US !GP
 3*0
 id 0 ID !GP
 Trait.Rider 2
 Tr 0 US !GP
 3*0
```

Rider 0 ID !GP

Appendix 7. Variance components for random effects of additive genetics (heritability), permanent environment and rider as a proportion of the total phenotypic variance

Trait	DA	DI	DN	DP	SJA	SJI	SJN	SJP	XCA	XCI	XCN	XCP
Heritability	0.09	0.076	0.072	0.076	0.162	0.095	0.094	0.089	0.027	0.003	0.013	0.014
Permanent environment	0.209	0.198	0.169	0.17	0.049	0.074	0.075	0.062	0.066	0.087	0.085	0.071
Rider	0.291	0.244	0.223	0.232	0.113	0.113	0.091	0.084	0.12	0.118	0.117	0.092
Total phenotypic	0.987	1.036	0.993	1.011	0.947	0.908	0.863	0.841	0.956	0.974	0.960	0.800
Repeatability	0.298	0.273	0.241	0.246	0.211	0.169	0.169	0.15	0.093	0.091	0.098	0.085